INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Large deviations for a class of Markov processes modelling communication networks

Franck Delcoigne — Arnaud de La Fortelle

N° 4474

June 2002

THÈME 1

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--4474--FR+ENG

Large deviations for a class of Markov processes modelling communication networks

Franck Delcoigne* , Arnaud de La Fortelle†

Thème 1 — Réseaux et systèmes
Projet PREVAL

Rapport de recherche n° 4474 — June 2002 — 42 pages

Abstract: In this paper, we prove a sample path large deviation principle (LDP) for a rescaled process $n^{-1}Q_{nt}$, where Q_t is a multi-dimensional birth and death process describing the evolution of a communication network. In this setting, Q_t is the join number of documents on the set of routes $\mathcal{R}$ at time t. Documents to be transferred arrive on route $r \in \mathcal{R}$ as a Poisson process with rate λ_r and are transferred at rate $\mu_r \nu_r(x)$ where x represents the state of the network, μ_r^{-1} is the mean size of documents on route r and $\nu_r(x)$ is the bandwidth allocated to route r.

We describe a set of assumptions over the allocation ν under which the LDP holds. Since we want the "classical" allocations to verify these assumptions, the difficulty is to deal with weak properties. For example, $\nu_r(x)$ is assumed to be continuous on the set $\mathcal{D}_r = \{x : x_r > 0\}$ but may be discontinuous elsewhere. Several examples are provided including the max-min-fairness allocation, a classical one in the context of data networks.

Since the main object to work with is the local rate function, a great care has been devoted to its expression and its properties. It is expressed as the solution of a convex program from which many useful properties are derived. We believe that this kind of expression allows numerical computations.

Key-words: Large deviations, empirical generator, change of measure, contraction principle, entropy, bandwidth sharing, convex program.

* Postal address: EDF R&D - Dept MRI - 1, avenue du Général de Gaulle - 92141. Clamart cedex - FRANCE. This work was partly done at FT R&D in the team managed by James Roberts.

† Postal address: INRIA - Domaine de Voluceau, Rocquencourt - BP 105 - 78153 Le Chesnay Cedex - FRANCE. e-mail address: `Arnaud.De_La_Fortelle@inria.fr`. This research is supported by the French Ministère de l'Équipement.

Grandes déviations pour une classe de processus de Markov modélisant des réseaux de télécommunications

Résumé : Dans cet article, nous prouvons un principe de grandes déviations (PGD) trajectoriel pour un processus renormalisé $n^{-1}Q_{nt}$, où Q_t est un processus de vie et de mort multi-dimensionnel décrivant l'évolution d'un réseau de télécommunication constitué d'un ensemble de routes $\mathcal{R}$. Le vecteur Q_t représente le nombre de documents en transfert sur chacune des routes $r \in \mathcal{R}$ au temps t. Les documents arrivent sur la route $r \in \mathcal{R}$ selon un processus de Poisson de paramètre λ_r et sont transférés à un taux $\mu_r \nu_r(x)$ où x représente l'état du réseau, μ_r^{-1} est la taille moyenne des documents circulant sur la route r et $\nu_r(x)$ est la bande passante allouée à la route r.

Nous donnons des conditions sur l'allocation ν sous lesquelles le PGD est valide. Nous souhaitons inclure dans le champ d'application la plupart des allocations "classiques" et c'est pourquoi ces hypothèses sont assez faibles : par exemple $\nu_r(x)$ est supposée continue sur l'ensemble $\mathcal{D}_r = \{x : x_r > 0\}$ mais peut être discontinue ailleurs. Plusieurs exemples sont analysés, comme un réseau de topologie quelconque sous l'allocation max-min-équité qui est classique dans le contexte des réseaux de données.

Il apparaît que, dans ce contexte, la fonctionnelle *locale* est l'objet central : nous nous sommes donc attachés à l'étudier finement. Elle s'exprime comme solution d'un programme convexe ce qui permet d'en déduire un certain nombre de propriétés. Nous pensons que ce type d'expression permet de poursuivre l'analyse par des méthodes numériques.

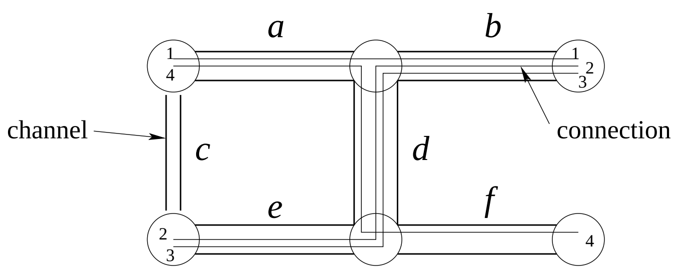
Mots-clés : Grandes déviations, générateur empirique, changement de mesure, principe de contraction, entropie, partage de bande passante, programme convexe

1 Introduction

The model Consider a network consisting of a set of links denoted $\mathcal{L}$. A route is a subset of links and the set of routes will be denoted by $\mathcal{R}$. Denote by $q_r(t)$ (resp. $q_l(t)$) the number of calls on route r (resp. the number of calls involving link l) at time t. In the sequel, r and l will represent respectively a route and a link. Each link has a capacity or a bandwidth equal to C_l (expressed for example in bits per seconde in the context of communication networks). Note that $q_l(t) = \sum_{r \ni l} q_r(t)$. Then $Q(t,x) = (q_r(t), r \in \mathcal{R})$ represents the state of the network at time t when it starts initially from state x. For the sake of simplicity, we shall sometimes omit x or t when they do not play a role.

Documents to be transferred arrive on route r according to a Poisson process of rate λ_r. The size of a document (expressed in bits) on route r is supposed to be exponentially distributed with parameter μ_r. Each document on route r is allocated a portion $\nu_r(x)/x_r$ of the bandwidth when the state of the network is x. Hence a document on route r is transferred at rate $\mu_r(x) \stackrel{\text{def}}{=} \mu_r \nu_r(x)$. In the sequel, it will be assumed that $\nu_r(x)$ satisfies the following faisability condition for all $x \in \mathbb{R}_+^{\mathcal{R}}$

$$\sum_{r \ni l} \nu_r(x) \leq C_l, \qquad \forall l \in \mathcal{L}. \tag{1.1}$$



routes r	connections per route q_r
$r_1 = (ab)$	$q_{r_1} = 1$
$r_2 = (bde)$	$q_{r_2} = 2$
$r_3 = (adf)$	$q_{r_3} = 1$

connections per link q_l	
$q_a = 2$	$q_d = 3$
$q_b = 3$	$q_e = 2$
$q_c = 0$	$q_f = 1$

Figure 1: Network model.

In our settings, $Q = \{Q(t, x_0),\ t \geq 0\}$ is a Markov process with generator R such that

$$Rf(x) = \sum_{y \in \mathbb{Z}_+^{\mathcal{R}}} q(x,y)\big(f(y) - f(x)\big), \qquad \forall x \in \mathbb{Z}_+^{\mathcal{R}},\ \forall f \in \mathcal{B}\big(\mathbb{Z}_+^{\mathcal{R}}\big),$$

$$\text{where} \quad q(x,y) \stackrel{\text{def}}{=} \begin{cases} \lambda_r, & \text{if } y - x = e_r, \\ \mu_r(x), & \text{if } y - x = -e_r \text{ and } x_r > 0, \\ 0, & \text{otherwise,} \end{cases}$$

Result In this paper, we aim at deriving a sample path LDP for the rescaled process

$$Q_x^n \stackrel{\text{def}}{=} \left\{ \frac{1}{n} Q(nt, [nx]),\ t \geq 0 \right\}.$$

Definition 1.1 (LDP) *The sequence $\{Q_x^n,\ n \geq 1\}$ satisfies a LDP in $D\big([0,T], \mathbb{R}_+^{\mathcal{R}}\big)$ with good rate function I_T if, for every $T > 0$, $x \in \mathbb{R}_+^{\mathcal{R}}$,*

(i) For any compact set $C \subset \mathbb{R}_+^{\mathcal{R}}$, $\bigcup_{x \in C} \Phi_x(K)$ is compact in $\mathcal{C}([0,T], \mathbb{R}_+^{\mathcal{R}})$ where Φ_x stands for the level sets of I_T:

$$\Phi_x(K) \stackrel{\text{def}}{=} \{\varphi \in D([0,T], \mathbb{R}_+^{\mathcal{R}}) : I_T(\varphi) \leq K,\ \varphi(0) = x\}.$$

(ii) for each closed set F of $D\big([0,T], \mathbb{R}_+^{\mathcal{R}}\big)$,

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left[Q_x^n \in F\right] \leq -\inf\{I_T(\phi),\ \phi \in F, \phi(0) = x\};$$

(iii) for each open set O of $D\big([0,T], \mathbb{R}_+^{\mathcal{R}}\big)$,

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left[Q_x^n \in O\right] \geq -\inf\{I_T(\phi),\ \phi \in O, \phi(0) = x\}.$$

Theorem 1.2 (LDP) *Under Assumptions (H1)-(H6), the sequence $\{Q_x^n,\ n \geq 1\}$ satisfies a LDP in $D\big([0,T], \mathbb{R}_+^{\mathcal{R}}\big)$ with the good rate function I_T defined by*

$$I_T(\varphi) \stackrel{\text{def}}{=} \begin{cases} \displaystyle\int_0^T L\left(\varphi(t), \dot{\varphi}(t)\right) dt, & \text{if } \varphi \text{ is absolutely continuous,} \\ +\infty, & \text{otherwise.} \end{cases} \tag{1.2}$$

The local rate function $L(x, D)$ is explicitely computed in Theorem 1.7.

Remark : I_T is defined by all the values $L(x, D)$, $x \in \mathbb{R}_+^{\mathcal{R}}$ and $D \in \mathbb{R}^{\Lambda(x)}$ (i.e. such that $D_r = 0$, $\forall r \in \Lambda^c(x)$). Indeed, assume that for some t, $\varphi_r(t) = 0$ and $\dot{\varphi}_r(t)$ exists. Since $\varphi_r(t) \le \varphi_r(s)$ for all s, this implies $\dot{\varphi}_r(t) = 0$. Moreover, φ being absolutely continuous, $\dot{\varphi}_r(t)$ exists for almost all t.

♦♦♦

Our main concern will be to identify the rate function I_T. Moreover, one of our contribution will be to describe a suitable set of conditions in the context of bandwith sharing under which one can express the rate function as the solution of a convex optimisation problem which may be completely explicit in some cases. As it will emerge in Section 2, the assumptions are satisfied for a variety of examples including the max-min-fairness allocation, a classical one in the context of data networks [2]. Although some literature has been devoted to the study of such systems and of their ergodicity conditions [3, 10, 18], it appears very hard to compute quantities of interest. In this regard, the calculation of the rate function could be a preliminary step in order to perform importance sampling or to compute the exponential decay of a route in stationary regime. All this program falls into the framework of LDP for Markov processes with discontinuous statistics, i.e those for which the coefficients of their generator are not spatially continuous.

Previous work It seems that one of the first paper dealing with such processes is [9], where large deviations problems for Jackson networks were investigated using partial differential equations techniques. Quite recently, the LDP for a large class of Markov processes with discontinuous statistics has been proved in [7]. Roughly speaking, the authors of [7] express the logarithm of large deviation probabilities as the minimal cost of some stochastic optimal-control problem, and the limit of the optimal cost is shown to exist by means of a sub-additivity argument. However, the rate function is not explicit. Note that in [8], an explicit upper bound of large deviations involving Legendre transforms is proved. While for Jackson networks and some processor sharing models this bound is tight [1], in general the problem of the lower bound remains open. However, in the present setting this upper bound is tight as it is proved in Theorem 1.7 and Section 4.4.

Untill now, the identification of the rate function has been carried out in some particular cases and usually for low dimensional systems. In this respect, general results were obtained in [6, 13] where the LDP has been established for random walks whose generator has a discontinuity along an hyperplane. These results are applied in [13] to compute the exponential decay of the stationary distribution of ergodic random walks in $\mathbb{Z}_+^2$. Nevertheless, in such examples, there are at most two boundaries with codimension one or two where discontinuity arise. Ultimately, the identification of the rate function governing the LDP for Jackson networks has been carried out in [1, 11]. It is worth noting that quite recently, the calculation of the rate function has been done for a large class of Markov additive process in [12]. The rate function is expressed in terms of the spectral radius of a class of infinite dimensional matrices associated to the original process. The setting of the present paper is in some sense more restrictive but the result more explicit. With respect to [1], we do not require the regularity of the Skorokhod map associated to the process, which is not proven to our knowledge. Instead of, we rely on the use of empirical measures and on the ergodicity conditions of a class of Markov processes associated to the original one. It appears also that the use of empirical measures is a constructive method allowing a careful description of how large deviations events occur.

Notation

- $r \ni l$ means "route r passes through link l" and $l \in r$ is synonymous ("l is used by r");
- For any set A, A^c will denote its complementary and $\mathbb{1}_{\{A\}}$ its indicator function;
- for any Banach space E, $\mathcal{B}(E)$, represents the set of bounded functions on E;
- $D([0,T],E)$ is the space of right continuous functions $[0,T] \to E$ with left limits, endowed with the Skorokhod metric d_d;
- $\mathcal{C}([0,T],E)$ is the space of continuous functions equipped with the metric of the uniform convergence d_c.

Definition 1.3 (Face) *For $x \in \mathbb{R}_+^{\mathcal{R}}$, the face $\Lambda(x)$ is defined by:*

$$\Lambda(x) \stackrel{\text{def}}{=} \{r \in \mathcal{R} : x_r > 0\}.$$

By an abuse of notation, we shall also call face Λ:

$$\left\{y \in \mathbb{R}_+^{\mathcal{R}} : y_r > 0,\ \forall r \in \Lambda, \quad \textit{and} \quad y_r = 0,\ \forall r \in \Lambda^c\right\}. \tag{1.3}$$

Assumptions In order to get a LDP for Q_x^n, it is natural to assume the property of homogeneity of the allocation, i.e:

Assumption (H1) (Homogeneity) *The function $\nu_r(x)$ is homogeneous:*

$$\nu_r(\alpha x) = \nu_r(x), \qquad \forall x \in \mathbb{R}_+^{\mathcal{R}},\ \forall \alpha > 0.$$

One can wonder whether the allocation, defined on $\mathbb{Z}_+^{\mathcal{R}}$ can be extended "naturally" to $\mathbb{R}_+^{\mathcal{R}}$. Obviously this is not a general property, but one can see that Assumption (H1) is valid for strictly positive integers α. Hence, defining $\nu_r(x/\alpha) \stackrel{\text{def}}{=} \nu_r(x)$, one extends "naturally" ν_r to $\mathbb{Q}_+^{\mathcal{R}}$ and with a mild continuity to $\mathbb{R}_+^{\mathcal{R}}$. Example are shown in Section 2.

Besides, in the context of bandwith sharing we will focus on allocations such that adding a connection anywhere in the system does not change drastically the bandwith allocated to route r when it is saturated. Using the homogeneity of the allocation, this can be stated as a property of continuity of the allocation *outside the boundaries.*

Assumption (H2) (Continuity) *The allocation $\nu_r(x)$ is continuous on the domain $\mathcal{D}_r \stackrel{\text{def}}{=} \{x : x_r > 0\}$.*

Assumption (H2) will allow one to compute in an easy way the cost in (1.7) for the components indexed by $\Lambda(x)$ to follow the prescribed growth rate D. Indeed, as a straight consequence of Assumptions (H1) and (H2), we derive the following property of $\nu_r(x)$:

Proposition 1.4 *Fix $x \in \mathbb{R}_+^{\mathcal{R}}$. Then,*

$$\lim_{n\to\infty} \nu_r(nx_{\Lambda(x)}, y) = \nu_r(x), \qquad \forall r \in \Lambda(x),\ \forall y \in \mathbb{R}_+^{\Lambda^c(x)}.$$

Moreover, the limit is uniform w.r.t. x and y in the following sense:

$$\lim_{\epsilon\to 0}\lim_{n\to\infty} \sup_{\substack{|y|<\epsilon n\\ |x'-x|<\epsilon}} |\nu_r(nx'_{\Lambda(x)}, y) - \nu_r(nx_{\Lambda(x)}, y)| = 0, \qquad \forall r \in \Lambda(x).$$

Proof : By homogeneity (H1),

$$\nu_r(nx'_{\Lambda(x)}, y) - \nu_r(nx_{\Lambda(x)}, y) = \nu_r(x'_{\Lambda(x)}, y/n) - \nu_r(x_{\Lambda(x)}, y/n).$$

Hence, using the continuity (H2)

$$\begin{aligned} &\lim_{\epsilon\to 0}\lim_{n\to\infty} \sup_{\substack{|y|<\epsilon n\\ |x'-x|<\epsilon}} |\nu_r(nx'_{\Lambda(x)}, y) - \nu_r(nx_{\Lambda(x)}, y)| \\ = &\lim_{\epsilon\to 0} \sup_{\substack{|y|<\epsilon\\ |x'-x|<\epsilon}} |\nu_r(x'_{\Lambda(x)}, y) - \nu_r(x_{\Lambda(x)}, y)| = 0. \end{aligned}$$

The proof is concluded. ∎

Now, we turn to the computation of the cost for the components belonging to $\Lambda^c(x)$ to stay in a neighborhood of 0. First, we assume that the allocation on the routes belonging to $\Lambda^c(x)$ behaves well when the components belonging to $\Lambda(x)$ are saturated.

Assumption (H3) (Behaviour on boundaries) *Fix $x \in \mathbb{R}_+^{\mathcal{R}}$.*
Then the following limit exists:

$$\nu_r^x(y) \stackrel{\text{def}}{=} \lim_{n\to\infty} \nu_r(nx_{\Lambda(x)}, y), \qquad \forall r \in \Lambda^c(x),\ \forall y \in \mathbb{R}_+^{\Lambda^c(x)}.$$

Moreover, the limit is uniform w.r.t. $y \in \mathbb{R}_+^{\Lambda^c(x)}$ in the following sense:

$$\lim_{\epsilon\to 0} \lim_{n\to\infty} \sup_{|z-x|<\epsilon, |y|<\epsilon n} |\nu_r^x(y) - \nu_r(nz_{\Lambda(x)}, y)| = 0, \qquad \forall r \in \Lambda^c(x).$$

At the moment, we have only assumed rather weak conditions in the context of bandwith sharing. We turn now to properties of ergodicity of a family of Markov processes associated to Q. The following notion will be at the heart of the present paper in proving the lower bound.

Localized processes Let $x \in \mathbb{R}_+^{\mathcal{R}}$. We now introduce a class of Markov processes which will be useful in dealing with the large deviations events quantified in Theorem 1.7 when the routes belonging to $\Lambda(x)$ are saturated[1] in proportion of x. First, let us describe the behaviour of the allocation in such situation. For routes $r \in \Lambda^c(x)$, it is described by Assumption (H3) by which $\nu_r^x(z)$ depends on $z \in \mathbb{Z}_+^{\Lambda^c(x)} \times \mathbb{Z}^{\Lambda(x)}$ only through the components belonging to $\Lambda^c(x)$, which are indeed all positive. For the routes pertaining to $\Lambda(x)$, by Proposition 1.4, the allocation is $\nu_r(x)$. Then define

$$\nu_r^x(z) \stackrel{\text{def}}{=} \begin{cases} \nu_r(x), & \text{if } r \in \Lambda(x), \\ \nu_r^x(z_{\Lambda^c(x)}), & \text{if } r \in \Lambda^c(x), \end{cases} \qquad \forall z \in \mathbb{Z}_+^{\Lambda^c(x)} \times \mathbb{Z}^{\Lambda(x)}. \tag{1.4}$$

Then define the intensity $\tilde{\mu}_r^x(z) \stackrel{\text{def}}{=} \tilde{\mu}_r \nu_r^x(z)$.

Now, $\tilde{Q}^x = \{\tilde{Q}^x(t, z_0),\ t \geq 0\}$ is precisely defined as the Markov process starting at z_0 with values in $\mathbb{Z}_+^{\Lambda^c(x)} \times \mathbb{Z}^{\Lambda(x)}$ and with generator $\tilde{R}^x$ such that for $f \in \mathcal{B}\big(\mathbb{Z}_+^{\Lambda^c(x)} \times$

[1] It describes the behaviour of the process Q *around* nx when n is large. For this reason, we shall call it *localized process*.

$\mathbb{Z}^{\Lambda(x)}\big)$,

$$\begin{aligned} \tilde{R}^x f(z) &= \sum_z \tilde{q}^x(z,z')\big(f(z')-f(z)\big), \qquad \forall z \in \mathbb{Z}_+^{\Lambda^c(x)} \times \mathbb{Z}^{\Lambda(x)}, \\ \text{where} \qquad \tilde{q}^x(z,z') &\stackrel{\text{def}}{=} \begin{cases} \tilde{\lambda}_r, & \text{if } z'-z = e_r, \\ \tilde{\mu}_r^x(z), & \text{if } z'-z = -e_r \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

Let us depict briefly the evolution of $\tilde{Q}^x$. The components belonging to $\Lambda(x)$ behave independently from each others and from the rest of the system as simple birth and death processes with birth rate $\tilde{\lambda}_r$ and death rate $\tilde{\mu_r}\nu_r(x)$. The components belonging to $\Lambda^c(x)$ behaves as a Markov process as the one presently studied but with $|\Lambda^c(x)|$ components (in other words with a new set of routes $\mathcal{R}^x = \Lambda^c(x)$) and allocation $\nu_r^x(.)$. Moreover, roughly speaking the available residual bandwidths for these routes are given by the numbers

$$C_l(x) \stackrel{\text{def}}{=} C_l - \sum_{r\in l,\ r\in\Lambda(x)} \nu_r(x), \qquad \forall l \in \mathcal{L}. \tag{1.5}$$

In fact, using (1.1), proposition 1.4 and the definition of $\nu_r^x(y)$ in (H3), one sees that $\nu_r^x(y)$ for $r \in \Lambda^c(x)$ satisfies the faisability condition

$$\sum_{r\ni l, r\in\Lambda^c(x)} \nu_r(x) \le C_l(x), \qquad \forall l \in \mathcal{L}. \tag{1.6}$$

In the sequel, the following partition of the routes belonging to $\Lambda^c(x)$ will be needed:

$$\begin{aligned} \Lambda_1(x) &= \{r \in \Lambda^c(x) : \exists l \in r,\ C_l(x) = 0\}, \\ \Lambda_2(x) &= \Lambda^c(x) \setminus \Lambda_1(x). \end{aligned}$$

$\Lambda_1(x)$ represents the set of non-saturated routes for which there is no residual bandwidth. We shall also assume that the allocation is not null on routes having some residual bandwidth available.

Assumption (H4) (Positivity) *For all $x \in \mathbb{R}_+^{\mathcal{R}}$ and $r \in \Lambda_2(x)$, it is assumed that $\nu_r^x(y) > 0$ when $y_r > 0$ for all $y \in \mathbb{R}_+^{\Lambda^c(x)}$.*

Remark : Taking $x = 0$, (H4) implies that $\nu_r(y) > 0$ when $y_r > 0$ for all $y \in \mathbb{R}_+^{\mathcal{R}}$.

♦♦♦

As a convention, if for some $r \in \Lambda^c(x)$, $\tilde{\lambda}_r = 0$, then we will take

$$\tilde{Q}_r^x(t, z_0) = 0, \ \forall t \geq 0.$$

For instance, only localized processes such that $\tilde{\lambda}_r = 0$ for $r \in \Lambda_1(x)$ will be considered and so assume that $\tilde{\lambda}_r = 0$ for $r \in \Lambda_1(x)$.

Remark : This is a slight modification of the process. In the following, only localized processes starting from $\lfloor nx \rfloor$ are used. In this case, clearly,

$$\tilde{Q}_r^x(t, \lfloor nx \rfloor) = \tilde{Q}_r^x(0, \lfloor nx \rfloor) = 0, \ \forall r \in \Lambda^c(x) \text{ such that } \tilde{\lambda}_r = 0.$$

and there is no modification. Nevertheless, this convention allows to define in a suitable way irreducibility and ergodicity properties of $\tilde{Q}_r^x$.

♦♦♦

Definition 1.5 *Denote $\tilde{Q}^x_{\Lambda^c(x)}$ the process defined by*

$$\tilde{Q}^x_{\Lambda^c(x)} \stackrel{\text{def}}{=} \left\{ \left(\tilde{Q}_r(t, z_0), \ r \in \Lambda_2(x) \ \textit{with} \ \tilde{\lambda}_r > 0\right), \ t \geq 0 \right\}.$$

Remark : The preceding convention consists simply in dropping the components belonging to $\Lambda^c(x)$ that do not play any role when dealing with ergodicity properties of the components of $\tilde{Q}^x$ belonging to $\Lambda^c(x)$.

♦♦♦

Using the definition of the generator $\tilde{R}^x$ and the modification of the process $\tilde{Q}^x$, $\tilde{Q}^x_{\Lambda^c(x)}$ is a Markov process. Besides, it is irreducible by the definition of $\tilde{\mu}_r^x(y)$ and by (H4).

Definition 1.6 *$\tilde{Q}^x$ is said ergodic if $\tilde{Q}^x_{\Lambda^c(x)}$ is.*

We now come to the main assumption which will allow one to express the cost in (1.7) for the components indexed by $\Lambda^c(x)$ to stay in a neighborhood of 0 as the solution of a convex program. This assumption will involve the conditions of ergodicity of localized processes.

Assumption (H5) (Conditions of ergodicity) *Fix $x \in \mathbb{R}_+^{\mathcal{R}}$. The localized process $\tilde{Q}^x$ is ergodic if, and only if,*

$$\sum_{r \in \Lambda_2(x),\ r \ni l} \frac{\tilde{\lambda}_r}{\tilde{\mu}_r} < C_l(x), \qquad \forall l \in \mathcal{L}.$$

These conditions will be called the usual conditions of ergodicity.

Note that taking $x = 0$ this assumption implies that the conditions of ergodicity for the network read

$$\sum_{r \ni l} \frac{\lambda_r}{\mu_r} < C_l, \qquad \forall l \in \mathcal{L}.$$

Owing to (1.1), these conditions are necessary. However, in many cases of interest, they are sufficient and Assumption (H5) is satisfied. Some examples are provided in the next section.

In order to establish a sample path LDP, one of the main task is to establish the local linear large deviation bounds of Theorem 1.7.

Theorem 1.7 *Let $x \in \mathbb{R}_+^{\mathcal{R}}$ and $D \in \mathbb{R}^{\Lambda(x)}$. Then, under Assumptions (H1)-(H5), writing $\lim_{\tau,\delta,\epsilon\to 0}$ for $\lim_{\tau\to 0}\lim_{\delta\to 0}\lim_{\epsilon\to 0}$,*

$$\begin{aligned}
&\lim_{\tau,\delta,\epsilon\to 0} \liminf_{n\to\infty} \frac{1}{n\tau} \inf_{|y-nx|<\epsilon n} \log \mathbb{P}\left[\sup_{t\in[0,n\tau]} |Q(t,y) - nx - Dt| < \delta n\right] \\
&= \lim_{\tau,\delta,\epsilon\to 0} \limsup_{n\to\infty} \frac{1}{n\tau} \sup_{|y-nx|<\epsilon n} \log \mathbb{P}\left[\sup_{t\in[0,n\tau]} |Q(t,y) - nx - Dt| < \delta n\right]
\end{aligned} \tag{1.7}$$

Moreover, if a face Λ and a growth rate $D \in \mathbb{R}^{\Lambda}$ are fixed, then the preceding limit in τ is uniform w.r.t. x in compact sets of Λ (see Definition 1.3). The common value of these limits is denoted by $-L(x,D)$ and

$$L(x,D) = \sum_{r\in\Lambda(x)} l\big(D_r \| \lambda_r, \mu_r \nu_r(x)\big) + \inf_{\eta\in V(x)} \sum_{r\in\Lambda^c(x)} l\big(0\|\lambda_r, \mu_r\eta_r\big), \tag{1.8}$$

$$\text{where} \qquad V(x) \stackrel{\text{def}}{=} \left\{\eta \in \mathbb{R}_+^{\Lambda^c(x)} : \sum_{r\in l,\ r\in\Lambda^c(x)} \eta_r \le C_l(x),\ \forall l \in \mathcal{L}\right\} \tag{1.9}$$

and where the positive quantity

$$l(D\|\lambda,\mu) \stackrel{\text{def}}{=} D \log\left(\frac{D + \sqrt{D^2 + 4\lambda\mu}}{2\lambda}\right) + \lambda + \mu - \sqrt{D^2 + 4\lambda\mu} \tag{1.10}$$

stands for the cost that a birth and death process with birth rate λ and death rate μ, follows the growth rate D (e. g. see [17]).

Remark : For $r \in \Lambda_1(x)$, the quantity $l\big(D_r \| \lambda_r, \mu_r \nu_r(x)\big)$ [or $l\big(D_r \| \lambda_r, \mu_r \eta_r\big)$] is easily determined, since there are hard constraints: $D_r = 0$ and $\nu_r(x) = 0$ [resp. $\eta_r = 0$], so that

$$l\big(D_r \| \lambda_r, \mu_r \nu_r(x)\big) = l\big(D_r \| \lambda_r, \mu_r \eta_r\big) = l(0 \| \lambda_r, 0) = \lambda_r$$

and so

$$L(x, D) = \sum_{r \in \Lambda(x)} l\big(D_r \| \lambda_r, \mu_r \nu_r(x)\big) + \sum_{r \in \Lambda_1(x)} \lambda_r + \inf \sum_{r \in \Lambda_2(x)} l\big(0 \| \lambda_r, \mu_r \eta_r\big)$$

where the infimum is taken over the set

$$\Big\{ \eta \in \mathbb{R}_+^{\Lambda_2(x)} : \sum_{r \in l,\ r \in \Lambda_2(x)} \eta_r \le C_l(x),\ \forall l \in \mathcal{L} \Big\}.$$

In conclusion, $V(x)$ and the corresponding infimum in (1.8) depend only on the values indexed by $\Lambda_2(x)$.

♦♦♦

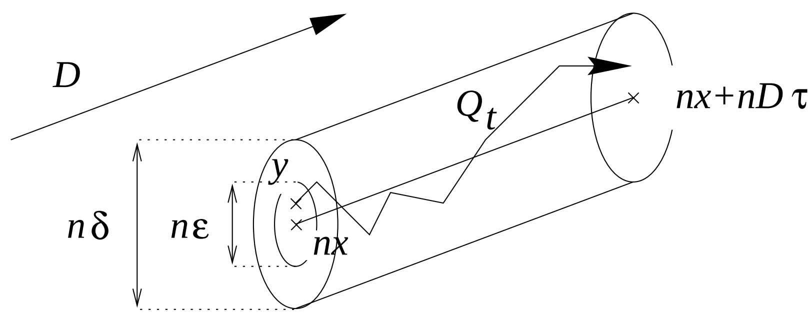


Figure 2: Structure of the local linear bounds of Theorem 1.7. $L(x, D)$ is the cost per unit time for the path $Q(t, y)$ (starting near nx) to stay in the neighborhood of $nx + Dt$ over the time $t \in [0, n\tau]$.

When proving the sample path LDP, one must extend the bounds obtained in Theorem 1.7 to bounds around linear paths (see [5, Appendix B]). The following property will be needed in order to deal with paths reaching a boundary (this is the *accessibility condition*).

Assumption (H6) (slow decrease near boundary) *There exists a constant* $C > 0$ *such that*

$$\nu_r(x) \geq C \frac{x_r}{\|x\|}, \qquad \forall r \in \mathcal{R},\ \forall x \in \mathbb{R}_+^{\mathcal{R}}.$$

Since in $\mathbb{R}^{\mathcal{R}}$ all norms are equivalent, this does not depend on the norm $\|.\|$.

Some comments about (H5) The main assumption is (H5). This assumption is rather natural in the context of Pareto efficient allocation. Let us recall that an allocation $\nu_r(x)$ is Pareto efficient if for all $x \in \mathbb{R}_+^{\mathcal{R}}$ and then all route $r \in \mathcal{R}$, there exists some link $l \in r$ where the capacity constraint C_l is attained. It means that in some sense no bandwith is wasted. Then in this setting, it is rather natural to guess that for a variety of Pareto efficient allocations, the conditions of ergodicity will be the usual ones. This is the case for the max-min-fairness and others described in Section 2.

2 Examples

The goal of this section is to show that Assumptions (H1)-(H6) are satisfied for a variety of examples. We start with a simple one.

2.1 Generalised processor sharing

In this case there is a single link $\{1\}$ with bandwidth C and the link is shared by several classes of customers indexed by $r \in \mathcal{R}$. A set of positive weights w_r are fixed. The allocation for the class i is given by

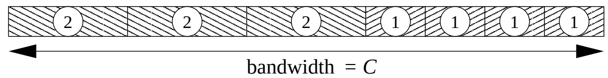
$$\nu_i(x) = \frac{w_i x_i}{\sum_{r \in \mathcal{R}} w_r x_r} C, \qquad \forall i \in \mathcal{R}.$$

Assumptions (H1)-(H3), (H4) and (H6) are easily seen to be satisfied. Moreover, for any $x \in \mathbb{R}_+^{\mathcal{R}}$, except for $x = 0$,

$$\nu_r^x(y) = 0, \qquad \forall r \in \Lambda^c(x),\ \forall y \in \mathbb{R}_+^{\Lambda^c(x)}.$$

Hence in this case $\Lambda_2(x) = \emptyset$ and Assumption (H5) is straightforward. Then

$$L(x, D) = \sum_{r \in \Lambda(x)} l\big(D_r \| \lambda_r, \mu_r \nu_r(x)\big) + \sum_{r \in \Lambda^c(x)} \lambda_r.$$



$$x_1 = 4 \quad w_1 = 1 \quad \nu_1(x) = 2C/5,$$
$$x_2 = 3 \quad w_2 = 2 \quad \nu_2(x) = 3C/5.$$

Figure 3: A processor sharing model with two classes.

2.2 The α-bandwidth allocation

We turn now to the general network described in the introduction. As proposed in [15], for α strictly positive and $\alpha \neq 1$, the allocation $\nu(x) = \big(\nu_r(x)\big)_{r\in\mathcal{R}}$ can be obtained as solution of the optimization problem under bandwidth constraints:

$$\max_{\nu} \left\{ \sum_{r\in\mathcal{R}} x_r^{\alpha} \frac{\nu_r^{1-\alpha}}{1-\alpha} \quad \text{where} \quad \sum_{r\ni l} \nu_r \leq C_l,\ \forall l \in \mathcal{L} \right\} \tag{2.1}$$

This is called the α-bandwidth allocation. Due to the concavity of the objective function (the first sum in (2.1)), the allocation $\nu_r(x)$ is uniquely defined. Assumptions (H1), (H2) follow by direct inspection.

Assumption (H6) Solving the optimization problem (2.1) yields:

$$\nu_r = x_r \Big(\sum_{l\in r} \beta_l \Big)^{-1/\alpha}$$

where β_l is the Lagrange multiplier associated to constraint $\sum_{l\in r} \nu_r \leq C_l$. The multipliers are positive; null if the constraint is not saturated. If the constraint l is saturated:

$$C_l = \sum_{r\ni l} \nu_r = \sum_{r\ni l} x_r \left(\sum_{l'\in r} \beta_{l'} \right)^{-1/\alpha} \leq \sum_{r\ni l} x_r \beta_l^{-1/\alpha} \leq \|x\|_1 \beta_l^{-1/\alpha},$$

where $\|x\|_1 \stackrel{\text{def}}{=} \sum_r x_r$. Then we deduce

$$\beta_l \leq C_l^{-\alpha} \|x\|_1^{\alpha}.$$

This inequality also holds for non-saturated links l, so that

$$\nu_r \geq x_r \left(\sum_{l \in r} C_l^{-\alpha} \|x\|_1^\alpha \right)^{-1/\alpha} \geq \frac{x_r}{\|x\|_1} |\mathcal{L}|^{-1/\alpha} \inf_{l \in \mathcal{L}} C_l,$$

where $|\mathcal{L}|$ is the cardinality of the set of links $\mathcal{L}$. The assumption is proved.

Assumption (H3) Let us turn to (H3). By Proposition 1.4 (which depends only on (H1) and (H2),

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} \sup_{|y| < \epsilon n} |\nu_r(nx_{\Lambda(x)}, y_{\Lambda^c(x)}) - \nu_r(x)| = 0, \qquad \forall r \in \Lambda(x).$$

So, for $\delta > 0$, there exist β and N such that for $\epsilon < \beta$ and $n > N$, the quantities $\nu_r(nx_{\Lambda(x)}, y_{\Lambda^c(x)})$, for $r \in \Lambda^c(x)$, are less than the solution of the optimization problem

$$\max_{\nu} \left\{ \sum_{r \in \Lambda^c(x)} x_r^\alpha \frac{\nu_r^{1-\alpha}}{1-\alpha} \quad \text{where} \quad \sum_{r \in \Lambda^c(x), r \ni l} \nu_r \leq C_l(x) + \delta, \ \forall l \in \mathcal{L} \right\} \tag{2.2}$$

and larger than the solution of the same optimization problem (2.2) with $C_l(x) + \delta$ replaced by $C_l(x) - \delta$. Letting δ tends to 0, (H3) is proved using the continuity of the objective function w.r.t ν at the point which is the solution of the optimisation problem, the continuity at this point following from (H6).

Assumptions (H5) and (H4) The preceding discussion ensures that $\tilde{Q}^x_{\Lambda^c(x)}$ depicts the evolution of a network of the type presently studied where:

(i) The set of routes is all $r \in \Lambda_2(x)$ with $\tilde{\lambda}_r > 0$.

(ii) The capacity on link l is equal to $C_l(x)$.

(iii) For the routes belonging to $\Lambda_2(x)$, the allocation is exactly the α-bandwidth allocation.

Since, as proved in [3] any network with the α-bandwidth allocation is ergodic if, and only if the usual conditions of ergodicity are satisfied, Assumption (H5) then holds. Moreover (H4) follows directly from (H6) and the preceding discussion: the localized allocation is itself an α-bandwidth allocation thus verifying (H6).

2.3 The max-min-fairness allocation

Let α of the α-bandwidth allocation tends to ∞. Then according to [15], one obtains the well known max-min fairness allocation [2], that is the fairest allocation. It is known that the numbers $\nu_r(x)/x_r$, the allocation dedicated to a given request on route r, can be obtained by means of the following "filling" procedure : Starting from an empty network, increase at the same speed the bandwidth allocated to each request until the capacity of some link is reached; then the allocation of the requests on routes using this link are frozen; one continues the procedure for the remaining flows and so on.

According to the filling procedure, (H1) and (H2) are satisfied. Moreover, it can be proved again by the filling procedure that (H3), (H4) and (H6) are satisfied and for $r \in \Lambda^c(x)$, $\nu_r^y(x)$ is the max-min-fairness allocation. Hence $\tilde{Q}^x_{\Lambda^c(x)}$ depicts a network where the allocation is the max-min-fairness and where the bandwith of link l is $C_l(x)$. Since by [18], any network with the max-min-fairness allocation is ergodic if, and only if, the usual conditions of ergodicity are satisfied, (H5) is then true.

2.4 The min policy

Although there exists some algorithm to compute the max-min-fairness allocation, it is not explicit. In order to get a more tractable model, the min policy was investigated in [10]. The allocation on route r is then given by

$$\nu_r(x) = x_r \min_{l \in r} \frac{C_l}{x_l}, \qquad \forall r \in \mathcal{R}.$$

It was proved in [10] that a network under the max min fairness allocation is stochastically smaller than the same network with the min policy. (H1), (H2), (H4) are straightforward. Moreover,

$$\Lambda_1(x) = \{r \in \mathcal{R} : \quad \exists l \in \mathcal{L}, r \ni l, x_l > 0\}.$$

Besides, for all $l \in \mathcal{L}$ such that $l \in r \in \Lambda_2(x)$ then $C_l(x) = C_l$. Moreover, in this context for $r \in \Lambda_2(x)$ the residual allocation $\nu_r^x(y)$ is given by

$$\nu_r^x(y) = y_r \min_{l \in r} \frac{C_l}{\displaystyle\sum_{r' \in \Lambda_2(x),\ r' \ni l} x_{r'}}, \qquad \forall r \in \Lambda_2(x).$$

Then Assumption (H3) can be easily verified. Besides, using the notation of Assumption (H5), $\tilde{Q}^x_{\Lambda^c(x)}$ depicts the evolution of a network with the min policy and all routes $r \in \Lambda_2(x)$ such that $\tilde{\lambda}_r > 0$. It is known from [10] that any network with the min policy is ergodic if and only if the usual conditions of ergodicity are fulfilled hence Assumption (H5) is satisfied. Hence Theorem 1.7 can be applied. Moreover, when the initial network is ergodic, i.e. when

$$\sum_{r \ni l} \frac{\lambda_r}{\mu_r} < C_l, \qquad \forall l \in \mathcal{L},$$

it can be readily checked that $\tilde{Q}^x_{\Lambda^c(x)}$ is also ergodic, which yields

$$\inf_{\eta \in V(x)} \sum_{r \in \Lambda_2(x)} l\big(0 \| \lambda_r, \mu_r \eta_r\big) = 0.$$

Hence, the local rate function reduces to the explicit expression

$$L(x, D) = \sum_{r \in \Lambda(x) \cup \Lambda_1(x)} l\big(D_r \| \lambda_r, \mu_r \nu_r(x)\big).$$

3 Localized generators

3.1 Localized empirical generator

Take $x \in \mathbb{R}_+^{\mathcal{R}}$ and $D \in \mathbb{R}^{\Lambda(x)}$. We are interested in computing large deviations bounds of the form (1.7) (i.e. linear bounds as presented in Figure 2). In order to prove Theorem 1.7, we introduce a functional which allows one to measure how the different arrival rates should be modified in order that the rescaled process Q^n_x follows a prescribed growth rate D. Let us introduce the localized empirical generator at point x as well as suitable state spaces associated to this process.

Definition 3.1 (Localized empirical generators) *Let $x \in \mathbb{R}_+^{\mathcal{R}}$. Denote by*

- *$A_r(t)$, the number of arrivals on route $r \in \mathcal{R}$ till t;*
- *$\nu_{\Lambda_2(x)}$ the restriction of the allocation ν to $\Lambda_2(x)$);*
- *$\Gamma^x \stackrel{\text{def}}{=} \mathbb{R}_+^{\mathcal{R}} \times \mathbb{R}_+^{\Lambda_2(x)} \times \mathbb{R}^{\mathcal{R}}$, the set of localized empirical generators;*

- G_t^x, *the localized empirical generator at point* x,

$$G_t^x = \left(\frac{1}{t}A(t), \frac{1}{t}\int_0^t \nu_{\Lambda_2(x)}\big(Q(s)\big)ds, \frac{Q_t - Q_0}{t}\right) \in \Gamma^x.$$

- $\Gamma^x \stackrel{\text{def}}{=} \mathbb{R}_+^{\mathcal{R}} \times \mathbb{R}_+^{\Lambda_2(x)} \times \mathbb{R}^{\mathcal{R}}$, *the set of localized empirical generators. The set* Γ^x *is equipped with the distance* d *defined for all* $G, G' \in \Gamma^x$ *by*

$$d(G,G') \stackrel{\text{def}}{=} \sum_{r\in\mathcal{R}} |a_r - a'_r| \ + \sum_{r\in\Lambda_2(x)} |\eta_r - \eta'_r| + \sum_{r\in\mathcal{R}} |D_r - D'_r|.$$

An element $G \in \Gamma^x$ has three components: $G = (A, \eta, D)$ which are meaningful:

- $A = (a_r)$ represents the mean arrival rates a_r on route $r \in \mathcal{R}$, hence $a_r \geq 0$;
- $\eta = (\eta_r)$ represents the mean bandwidth η_r allocated to route $r \in \Lambda_2(x)$, hence $\eta_r \geq 0$.
- $D = (D_r)$ represents the mean growth rates D_r of the number of connections on route $r \in \mathcal{R}$.

Since it is difficult to analyze at first the behavior of $Q(t)$ as in (1.7), we shall first analyze the behavior of G_t^x (which is a more detailed object than Q_t) from which we will deduce the behavior of Q_t. This means that we shall focus on the probability that G_t^x stays in a neighborhood of a prescribed generator G or more precisely, we shall establish large deviation bounds for the event

$$E_{\tau,\delta,y}^{(n)}(x,G) \stackrel{\text{def}}{=} \left\{ G_{n\tau}^x \in B(G,\delta), \ \sup_{t\in[0,n\tau]} |Q(t,y) - nx - nDt| < \delta n \right\} \tag{3.1}$$

where $B(G,\delta)$ is the ball of center G and radius δ (within the metric space (Γ^x, d)).

3.2 Localized generator

The localized empirical generator G_t^x do not evolve actually in the whole space Γ^x. First, $a_r - D_r$ is the mean departure rate on route r, hence $a_r - D_r \geq 0$. Second, on $E_{\tau,\delta,y}^{(n)}(x,G)$, by Assumptions (H1) and (H2), the components $\nu_r(Q_t)$ pertaining to $\Lambda(x)$ are close to $\nu_r(x)$. Since the mean allocations η_r exactly satisfy the bandwidth constraints (1.1), it can be checked that they satisfy the local bandwidth constraints

$$\sum_{r\ni l, r\in\Lambda_2(x)} \eta_r \leq C_l(x), \qquad \forall l \in \mathcal{L}. \tag{3.2}$$

The second type of constraints comes with the setting of the problem. All prescribed generator G are not relevant. Due to the remark p. 5 (following Theorem 1.2), the generators with $D_r \neq 0$ for $r \in \Lambda^c(x)$ can be rejected of the analysis. All these constraints motivates the introduction of the following subspace of Γ^x.

The third type of constraint comes from the fact that there is a set of empirical generators $G \in \Gamma^x$ such that $E^{(n)}_{\tau,\delta,y}(x, G)$ cannot occur at a large deviations scale. As it will emerge, when $D \in \mathbb{R}^{\Lambda(x)}$ this is the case if $\eta_r = 0$ and $a_r > 0$ for some $r \in \Lambda_2(x)$ or if $a_r > 0$ for some $r \in \Lambda_1(x)$. Both cases mean that there are departures (i.e. connections achieving their service) when the bandwidth allocation is null: obviously this event is highly improbable. All these constraints motivates the introduction of the following subspace of Γ^x.

Definition 3.2 *The set $\mathcal{G}^x \subset \Gamma^x$ denotes the set of localized generators $G = (A, \eta, D)$ such that:*

$$D_r = 0 \qquad \forall r \in \Lambda^c(x) \tag{3.3}$$

$$a_r = 0, \qquad \forall r \in \Lambda_1(x) \tag{3.4}$$

$$a_r = 0, \qquad \forall r \in \Lambda_2(x) \quad \text{with } \eta_r = 0, \tag{3.5}$$

$$\sum_{r \ni l, r \in \Lambda_2(x)} \eta_r < C_l(x), \qquad \forall l \in \mathcal{L} \text{ with } C_l(x) > 0, \tag{3.6}$$

$$a_r > 0 \text{ and } a_r - D_r > 0, \quad \forall r \in \Lambda(x). \tag{3.7}$$

In this setting $\overline{\mathcal{G}}^x$ will represent the closure of $\mathcal{G}^x$.

The generators $G \in \mathcal{G}^x$ can be seen as limit behavior for empirical generators G^x_t, when the time t and the departure point x are very large.

Now, we prove that in establishing large deviations bounds, it is sufficient to deal with generators belonging to $\mathcal{G}^x$. More precisely:

Lemma 3.3 *For $x \in \mathbb{R}_+^{\mathcal{R}}$ and $D \in \mathbb{R}^{\Lambda(x)}$, if $G = (A, \eta, D) \notin \overline{\mathcal{G}}^x$, then*

$$\lim_{\tau,\delta,\epsilon \to 0} \limsup_{n \to \infty} \frac{1}{n\tau} \sup_{|y - nx| < \epsilon n} \log \mathbb{P}\left[E^{(n)}_{\tau,\delta,y}(x, G)\right] = -\infty. \tag{3.8}$$

Proof : The proof is postponed to Appendix A. ■

3.3 Correspondence between localized Markov processes and localized generators

Let $G = (A, \eta, D) \in \overline{\mathcal{G}}^x$ be a localized generator. In order to prove a lower bound for the localized empirical generator, to G is associated a localized process $(\tilde{\lambda}_r, \tilde{\mu}_r, \nu_r(x))$. The parameters $\tilde{\lambda}_r$ and $\tilde{\mu}_r$ are defined in the following way:

$$
\begin{array}{llll}
\text{(i)} & \tilde{\lambda}_r = a_r, & \forall r \in \mathcal{R} & \\
\text{(ii)} & \tilde{\mu}_r \nu_r(x) = a_r - D_r, & \forall r \in \Lambda(x), & \\
\text{(iii)} & \tilde{\mu}_r \eta_r = a_r, & \forall r \in \Lambda^c(x), & \text{if } a_r > 0, \\
\text{(iv)} & \tilde{\mu}_r = \mu_r, & \forall r \in \Lambda^c(x), & \text{if } a_r = 0.
\end{array}
\tag{3.9}
$$

Then, the limiting behaviour of this localized process is given by the following important lemma.

Lemma 3.4 *Let $x \in \mathbb{R}_+^{\mathcal{R}}$, $G = (A, \eta, D) \in \mathcal{G}^x$ a localized empirical generator and denote by $\tilde{\mathbb{P}}$ the law of the corresponding localized Markov process $\tilde{Q}(t, \lfloor nx \rfloor)$. Then, for all $\tau > 0$,*

$$\lim_{\tau \to 0} \lim_{\delta \to 0} \liminf_{n \to \infty} \tilde{\mathbb{P}} \left[E_{\tau,\delta,x}^{(n)}(x, G) \cap \left\{ A_r(n\tau) = 0, \ \forall r \text{ with } a_r = 0 \right\} \right] = 1.$$

Proof : First note that $\tilde{\mathbb{P}}\left[A_r(n\tau) = 0, \ \forall r \text{ with } a_r = 0\right] = 1$ for all n and $\tau > 0$, since $\tilde{\lambda}_r = a_r$ by (3.9).

For the first term, since $G = (A, \eta, D) \in \mathcal{G}^x$,

$$\sum_{r \ni l, r \in \Lambda_2(x)} \frac{\tilde{\lambda}_r}{\tilde{\mu}_r} = \sum_{r \ni l, r \in \Lambda_2(x)} \eta_r < C_l(x), \ \forall l \in \mathcal{L}.$$

Hence the associated localized Markov process is ergodic by Assumption (H5). Then, $\tilde{\mathbb{P}}$ almost surely,

$$\lim_{n \to \infty} \frac{1}{n\tau} \int_0^{n\tau} \tilde{\mu}_r \nu_r^x\big(Q(s)\big)\, ds = \tilde{\lambda}_r, \qquad \forall r \in \Lambda_2(x). \tag{3.10}$$

For r with $a_r > 0$, this follows writing the balance equation for q_r and applying the ergodic theorem. For r with $a_r = 0$, this is follows from Remark 1. Hence, applying the law of large numbers for the arrival process yields

$$\lim_{n \to \infty} \left(\frac{1}{n\tau} A(n\tau), \frac{1}{n\tau} \int_0^{n\tau} \nu_{\Lambda_2(x)}^x\big(Q(s)\big)\, ds \right) = (A, \eta) \quad \text{a.s.} \tag{3.11}$$

There are three cases to consider in order to prove

$$\lim_{n\to\infty} \sup_{t\leq\tau} \left| \frac{1}{n} \int_0^{nt} \Big(\tilde\lambda_r - \tilde\mu_r \nu_r^x\big(Q(ns)\big)\Big) ds - D_r t \right| = 0, \qquad \forall r \in \mathcal{R}. \tag{3.12}$$

- For $r \in \Lambda(x)$, this is a consequence of (3.9) and of (1.4) which defines $\nu_r^x(y)$ for $r \in \Lambda(x)$;
- For $r \in \Lambda_1(x)$, this is immediate since by definition of $\mathcal{G}^x$ and (3.9), $\tilde\lambda_r = a_r = 0$, $D_r = 0$, and $\nu_r^x(y) = 0$ by definition of $\Lambda_1(x)$;
- For $r \in \Lambda_2(x)$, this follows from (3.10) and $D_r = 0$ since $G = (A, \eta, D) \in \mathcal{G}^x$.

On an other hand,

$$M_t^{(n)} \stackrel{\text{def}}{=} \frac{1}{n} q_r(nt) - \frac{1}{n} q_r(0) - \frac{1}{n} \int_0^{nt} \Big(\tilde\lambda_r - \tilde\mu_r \nu_r^x\big(Q(ns)\big)\Big) ds$$

is a martingale with Doob Meyer bracket

$$< M^{(n)} >_t = \frac{1}{n^2} \int_0^{nt} \Big(\tilde\lambda_r + \tilde\mu_r \nu_r^x\big(Q(ns)\big)\Big) ds$$

tending to 0 when n goes to infinity. So, Doob's inequality applied to $M_t^{(n)}$ and (3.12) yield

$$\lim_{\delta\to 0} \liminf_{n\to\infty} \tilde{\mathbb{P}} \left[\sup_{t\in[0,n\tau]} |Q(t, \lfloor nx \rfloor) - nx - Dt| < \delta n \right] = 1. \tag{3.13}$$

Using (3.13) and combining (3.11) with (H3), the proof is concluded. ■

4 The entropy and the rate function

4.1 Entropy

Definition 4.1 *Let $x \in \mathbb{R}_+^{\mathcal{R}}$, $R(x) = (\lambda_r, \mu_r, \nu_r(x))$ be the generator of the Markov process Q at x, $G = (A, \eta, D) \in \mathcal{G}^x$ be a localized generator and $\big(\tilde\lambda_r, \tilde\mu_r, \nu_r^x(y),\ y \in \mathbb{Z}^{\Lambda(x)} \times \mathbb{Z}_+^{\Lambda^c(x)}\big)$ its representation as a localized network model. The relative entropy of G with respect to $R(x)$ is*

$$H(G\|R(x)) = \sum_{r\in\mathcal{R}} I_p\big(\tilde\lambda_r \| \lambda_r\big) + \sum_{r\in\Lambda(x)} \nu_r(x) I_p\big(\tilde\mu_r\|\mu_r\big) + \sum_{r\in\Lambda_2(x)} \eta_r I_p\big(\tilde\mu_r\|\mu_r\big)$$

where $I_p(\nu\|\lambda)$ *is the relative entropy of Poisson processes of intensities* ν *and* λ *defined by*

$$I_p(\nu\|\lambda) \stackrel{\text{def}}{=} \nu \log \frac{\nu}{\lambda} - \nu + \lambda \tag{4.1}$$

with the convention $\frac{0}{0} = 0$ *and* $0 \log 0 = 0$.

The entropy has an easy interpretation in terms of information theory: it can be defined as the *mean information gain per unit time.* $H(.\|R)$ is decomposed as the sum of the information gain for the arrivals $I_p(\tilde{\lambda}_r\|\lambda_r)$, the information gain for the service time $I_p(\tilde{\mu}_r\|\mu_r)$ multiplied by the speed of service $\nu_r(x)$ for $r \in \Lambda(x)$ or multiplied by η_r for $r \in \Lambda_2(x)$. Note that if $\eta_r = 0$ for some $r \in \Lambda_2(x)$, then the information gain is null:Indeed, in this case, by (3.9) the service rate is left unchanged. This is also the case for $r \in \Lambda_1(x)$.

Lemma 4.2 *For fixed* x, $H(.\|R(x))$ *is continuous on* $\overline{\mathcal{G}}^x$.

Proof : It is an easy consequence of the expression (4.1). ∎

4.2 The local rate function $L(x, D)$

Definition 4.3 *The local rate function* $L(x, D)$ *is defined by*

$$L(x, D) \stackrel{\text{def}}{=} \inf_{G \in f_x^{-1}(D)} H\big(G\|R(x)\big), \qquad \forall D \in \mathbb{R}^{\Lambda(x)}, \tag{4.2}$$

where $f_x : \mathcal{G}^x \mapsto \mathbb{R}^{\Lambda(x)}$ *is the projection* $f_x(G) = D$.

Performing the optimisation of $H\big((A, \eta, D)\|R(x)\big)$ first w.r.t. A and then to η yields the explicit representation (1.8) for $L(x, D)$. Indeed, note that $r \in \Lambda_1(x)$ implies $\nu_r(x) = 0$. Moreover, the function $l(D\|\lambda, \mu)$ defined in (1.10) is obtained by optimizing the entropy:

$$l(D\|\lambda, \mu) = \inf_{a,\eta} \big\{H(a\|\lambda) + H(a - D\|\mu\eta)\big\}.$$

Extension of $L(x, D)$: In Definition 4.3, $L(x, D)$ is only defined for D such that $D_r = 0$ for all $r \in \Lambda^c(x)$. This definition of $L(x, D)$ can be extended to all $D \in \mathbb{R}^{\mathcal{R}}$ (using (1.8), e.g.), but due to the remark p. 5, this is not relevant.

In the rest of this section, a number of useful properties of $L(x, D)$ are derived. We start with a useful lemma:

Lemma 4.4 *If a face Λ is fixed, then $C_l(x)$ is continuous w.r.t. $x \in \Lambda$.*

Proof : Take $x^{(n)} \to x$ with $\Lambda\big(x^{(n)}\big) = \Lambda(x)$ for all n. Using (H2), $\nu_r\big(x^{(n)}\big) \to \nu_r(x)$, for all $r \in \Lambda(x)$. So

$$\lim_{n\to\infty} C_l\big(x^{(n)}\big) = C_l - \lim_{n\to\infty} \sum_{r\ni l, r\in\Lambda(x)} \nu_r\big(x^{(n)}\big) = C_l(x)$$

hence the result. ■

Proposition 4.5 *The local rate function $L(x,D)$ possesses the following properties.*

(i) *It is positive, finite, strictly convex and continuous with respect to D such that $D_r \geq 0$ for all $r \in \Lambda(x)$.*

(ii) *let $\|D\| \stackrel{\text{def}}{=} \max_i |D_i|$; there exists $M \in \mathbb{R}$ such that,*

$$L(x,D) \geq \frac{1}{2}\|D\| \log \|D\|, \qquad \forall x \in \mathbb{R}_+^{\mathcal{R}},\ \forall \|D\| \geq M;$$

In particular, it has compact level sets w.r.t. D;

(iii) *for a fixed D and a prescribed face Λ, $L(x,D)$ is continuous for $x \in \Lambda$ (see equation (1.3));*

(iv) *$L(x,D)$ is jointly lower semi-continuous w.r.t. x and D.*

Proof : (i) and (ii) are immediate from the explicit representation (1.8) of $L(x,D)$. Let $x^{(n)} \to x$ with $\Lambda\big(x^{(n)}\big) = \Lambda(x)$ for all n. Then

$$\begin{aligned} L\big(x^{(n)}, D\big) &= \sum_{r\in\Lambda(x)} l\big(D_r \| \lambda_r, \mu_r \nu_r(x^{(n)})\big) \\ &\quad + \inf_{\eta\in V(x^{(n)})} \sum_{r\in\Lambda^c(x)} l\big(D_r\|\lambda_r, \mu_r\eta_r\big), \end{aligned}$$

where $V(x)$ is defined in (1.9).

Since $l(D\|\lambda,\mu)$ is continuous w.r.t. μ, since $C_l\big(x^{(n)}\big) \to C_l(x)$ by Lemma 4.4 and since $\nu_r\big(x^{(n)}\big) \to \nu_r(x)$ for all $r \in \Lambda(x)$ by Assumption (H2), one obtains

$$\lim_{n\to\infty} L\big(x^{(n)}, D\big) = L(x,D)$$

hence (iii).

Let $(x^{(n)}, D^{(n)}) \to (x, D)$. For n sufficiently large $\Lambda(x) \subset \Lambda\big(x^{(n)}\big)$. So,

$$\begin{aligned} L\big(x^{(n)}, D^{(n)}\big) &= \sum_{r\in\Lambda(x)} l\big(D_r^{(n)} \| \lambda_r, \mu_r\nu_r(x^{(n)})\big) \\ &\quad + \sum_{r\in\Lambda\left(x^{(n)}\right)\setminus\Lambda(x)} l\big(D_r^{(n)} \| \lambda_r, \mu_r\nu_r(x^{(n)})\big) \\ &\quad + \inf_{\nu} \sum_{r\in\Lambda^c(x^{(n)})} l\big(D_r^{(n)} \| \lambda_r, \mu_r\nu_r\big) \end{aligned}$$

where the infimum is taken over the set of $\nu \in \mathbb{R}_+^{\Lambda^c(x^{(n)})}$ verifying

$$\sum_{\substack{l\ni r \\ r\in\Lambda^c(x^{(n)})}} \nu_r \;+\; \sum_{\substack{l\ni r \\ r\in\Lambda(x^{(n)})\setminus\Lambda(x)}} \nu_r\big(x^{(n)}\big) \;\le\; C_l - \sum_{\substack{l\ni r \\ r\in\Lambda(x)}} \nu_r\big(x^{(n)}\big), \qquad \forall l \in \mathcal{L}.$$

Using $\Lambda^c(x) = \Lambda^c\big(x^{(n)}\big) \cup \big(\Lambda\big(x^{(n)}\big) \setminus \Lambda(x)\big)$,

$$\begin{aligned} L(x^{(n)}, D^{(n)}) &\ge \sum_{r\in\Lambda(x)} l\big(D_r^{(n)} \| \lambda_r, \mu_r\nu_r(x^{(n)})\big) \\ &\quad + \inf_{\nu} \sum_{r\in\Lambda^c(x)} l\big(D_r^{(n)} \| \lambda_r, \mu_r\nu_r\big) \end{aligned}$$

where the infimum is taken over the set of $\nu \in \mathbb{R}_+^{\Lambda^c(x)}$ verifying

$$\sum_{l\ni r, r\in\Lambda^c(x)} \nu_r \;\le\; C_l - \sum_{l\ni r, r\in\Lambda(x)} \nu_r\big(x^{(n)}\big), \quad \forall l \in \mathcal{L}.$$

$l(D\|\lambda, \mu)$ being continuous w.r.t. μ and D and $\nu_r\big(x^{(n)}\big) \to \nu_r(x)$ for all $r \in \Lambda(x)$ by assumption (H2), one obtains

$$\liminf_{n\to\infty} L(x^{(n)}, D^{(n)}) \ge L(x, D)$$

hence (iv). ■

4.3 The sample path rate function I_T

In this section, we verify that the rate function I_T, defined by (1.2), possesses the usual properties.

Proposition 4.6 *The rate function I_T possesses the following properties.*

(i) Assume $I_T(\varphi) \leq K$ for some K. Then, for all $\epsilon > 0$, there exists $\delta > 0$ independent of φ such that for any collection of non overlapping intervals $[t_j, t_{j+1}]$ in $[0,T]$ with $\sum_j t_{j+1} - t_j = \delta$,

$$\sum_j |\varphi(t_{j+1}) - \varphi(t_j)| \leq \epsilon;$$

(ii) $I_T(.)$ is lower semi-continuous in $\Big(D\big([0,T], \mathbb{R}_+^{\mathcal{R}}\big), d_d\Big)$;

(iii) for $C \subset \mathbb{R}_+^{\mathcal{R}}$ compact, $\bigcup_{x \in C} \Phi_x(K)$ is compact in $\mathcal{C}([0,T], \mathbb{R}_+^{\mathcal{R}})$;

(iv) consider an absolutely continuous function φ with $I_T(\varphi) < \infty$. Then, for all $\epsilon > 0$, there exists a piecewise linear function φ_ϵ such that:

(a) $d_c(\varphi_\epsilon, \varphi) \leq \epsilon$,

(b) $I_T(\varphi_\epsilon) \leq I_T(\varphi) + \epsilon$.

Proof : One proves (i) using Proposition 4.5 (ii) in a way similar to [17, Lemma 5.18].

In order to prove the lower semi-continuity of $I_T(.)$, (i) shows it is sufficient to consider sequences of absolutely continuous functions. Since on $\mathcal{C}\big([0,T], \mathbb{R}_+^{\mathcal{R}}\big)$, the metrics d_c and d_d are equivalent, one can use d_c. Now, using Proposition 4.5 (ii), the fact that $L(x, D)$ is lower semi-continuous in (x, D) and convex with respect to D by Proposition 4.5, (ii) is proved by means of [14, Theorem 3 of Section 9.1.4].

(iii) is a consequence of (i) and (ii) (see [17, Proposition 5.46].

The proof of (iv) is a simple adaptation of [4, Proposition 5.1 (iv)]. ■

Remark : The points (ii) and (iv) imply that for any absolutely continuous φ with $I_T(\varphi) < \infty$, there exists a sequence of piecewise linear functions $\{\varphi_n,\ n \geq 1\}$ satisfying

$$\lim_{n \to \infty} d_c(\varphi_n, \varphi) = 0 \quad \text{and} \quad \lim_{n \to \infty} I_T(\varphi_n) = I_T(\varphi).$$

♦♦♦

4.4 Equivalence of I_T and the upper bound of [8]

In this section, we recall the upper bound of [8] in our setting and show the identity with our bounds. Denote

$$H(x,\alpha) = \sum_{r\in\mathcal{R}} \lambda_r\Big(e^{\alpha_r}-1\Big)+\mu_r\nu_r(x)\Big(e^{-\alpha_r}-1\Big)$$

and

$$h(x,\alpha) = \lim_{\delta\to 0}\sup_{y:|y-x|\le\delta} H(y,\alpha).$$

Then, introduce the Legendre-Fenchel transform of $h(x,\alpha)$

$$\tilde{L}(x,D) = \sup_{\alpha}\Big(<\alpha, D> - h(x,\alpha)\Big).$$

and the functional

$$\tilde{I}_T(\varphi) \stackrel{\text{def}}{=} \begin{cases} \displaystyle\int_0^T \tilde{L}\left(\varphi(t),\dot\varphi(t)\right)dt, & \text{if } \varphi \text{ is absolutely continuous,}\\ +\infty, & \text{otherwise.}\end{cases} \tag{4.3}$$

Then, [8, Theorem 1.1] reads:

Theorem 4.7 *Let a compact set $C\subset\mathbb{R}_+^{\mathcal{R}}$ be given. Then for each closed set F of $D\big([0,T],\mathbb{R}_+^{\mathcal{R}}\big)$,*

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left[Q_x^n\in F\right] \le -\inf\{\tilde{I}_T(\phi),\ \phi\in F, \phi(0)=x\}$$

uniformly in $x\in C$.

Moreover, this theorem applies for each localized process but with $\nu_r^x()$ replacing $\nu_r()$ in the definition of $H(y,\alpha)$.

For all $x\in\mathbb{R}_+^{\Lambda^c(x)}$, denote

$$\begin{aligned} F^x &= \left\{\eta\in\mathbb{R}_+^{\Lambda^c(x)} : \sum_{l\ni r, r\in\Lambda^c(x)} \eta_r \le C_l(x)\right\}\\ \mathcal{V}^x &= \left\{\eta\in\mathbb{R}_+^{\Lambda^c(x)} : \exists D, D_r\ge 0, \forall r\in\Lambda^c(x), \eta_r=\nu_r^x(D), \forall r\in\Lambda^c(x)\right\}\end{aligned}$$

Proposition 4.8 *Assume (H5) is fulfilled. Then, for all* $x \in \mathbb{R}_+^{\Lambda^c(x)}$,

$$\overline{\mathcal{V}}^x = F^x.$$

Moreover, if (H1), (H2), (H5) are fulfilled, then

(*i*) $L(x,D) = \tilde{L}(x,D)$, $\forall D \in \mathbb{R}^{\Lambda(x)}$.

(*ii*) $I_T(\varphi) = \tilde{I}_T(\varphi)$, $\forall \varphi \in D\big([0,T], \mathbb{R}_+^{\mathcal{R}}\big)$.

Proof : In order to prove the first assertion, note that (1.6) implies

$$\overline{\mathcal{V}}^x \subset F^x.$$

Let us turn to the opposite inclusion. We apply Theorem 4.7 for the localized process at x. Using (H5), $\forall (\lambda_r/\mu_r, r \in \Lambda^c(x)) \in \dot{F}^x$,

$$\tilde{L}^x(0,0) = 0. \tag{4.4}$$

where $\tilde{L}^x(y,D)$ is the equivalent of $\tilde{L}(x,D)$ but applied to the localized process at x. But

$$\tilde{L}^x(0,0) \geq \inf_{\eta \in \mathcal{V}^x} -\left(\sum_{r \in \Lambda^c(x)} \lambda_r \Big(e^{\alpha_r} - 1\Big) + \mu_r \eta_r \Big(e^{-\alpha_r} - 1\Big) \right), \ \forall \alpha.$$

Taking $e^{\alpha_r} = \sqrt{\mu_r \eta_r / \lambda_r}$ for all r yields

$$\tilde{L}^x(0,0) \geq \inf_{\eta \in \mathcal{V}^x} \sum_{r \in \Lambda^c(x)} \Big(\sqrt{\lambda_r} - \sqrt{\mu_r \eta_r} \Big)^2.$$

Since this is true for all $(\lambda_r/\mu_r, r \in \Lambda^c(x)) \in \dot{F}^x$, (4.4) implies $\dot{F}^x \subset \overline{\mathcal{V}}^x$. Finally $\overline{\mathcal{V}}^x = F^x$ concluding the proof of the first assertion.

Let us turn to (i). Using (H1), (H2), (H3) gives

$$\begin{aligned} H(x,\alpha) &= \sum_{r \in \Lambda(x)} \lambda_r \Big(e^{\alpha_r} - 1\Big) + \mu_r \nu_r(x) \Big(e^{-\alpha_r} - 1\Big) \\ &+ \sup_{\{y : y_r \geq 0\}} \sum_{r \in \Lambda^c(x)} \lambda_r \Big(e^{\alpha_r} - 1\Big) + \mu_r \nu_r^x(y) \Big(e^{-\alpha_r} - 1\Big). \end{aligned}$$

Hence using (i), for $D \in \mathbb{R}^{\Lambda(x)}$

$$
\begin{aligned}
\tilde{L}(x,D) &= \sum_{r\in\Lambda(x)} l\big(D_r \| \lambda_r, \mu_r\nu_r(x)\big) \\
&+ \sup_{\alpha}\Bigg(-\sup_{\{y:y_r\geq 0\}} \sum_{r\in\Lambda^c(x)} \lambda_r\Big(e^{\alpha_r}-1\Big)+\mu_r\nu_r^x(y)\Big(e^{-\alpha_r}-1\Big)\Bigg) \\
&= \sum_{r\in\Lambda(x)} l\big(D_r \| \lambda_r, \mu_r\nu_r(x)\big) \\
&+ \sup_{\alpha}\Bigg(-\sup_{\{\eta\in F^x\}} \sum_{r\in\Lambda^c(x)} \lambda_r\Big(e^{\alpha_r}-1\Big)+\mu_r\eta_r\Big(e^{-\alpha_r}-1\Big)\Bigg) \\
&= L(x,D)
\end{aligned}
$$

where the last equality follows from the Min-Max principle, see [16, Corollary 37.3.2 p. 393].

(ii) is a straight consequence of (i) and Remark 1. ■

5 Large deviations bounds for empirical generators

5.1 An exponential change of measure

Consider a multi dimensional birth and death process with birth and death rates given by $(\tilde{\lambda}_r, \tilde{\mu}_r(y), y \in \mathbb{R}_+^{\mathcal{R}}, r \in \mathcal{R})$ such that $\mu_r(y) > 0$ when $\tilde{\mu}_r(y) > 0$. Let us describe how $\tilde{\mathbb{P}}$ can be obtained from $\mathbb{P}$. For, denote by

- N_t, the number of jumps of the process till t.
- $Q(k) = \{Q_r(k),\ r \in \mathcal{R}\}$, the embedded Markov chain at time $k \in \mathbb{N}$. We shall distinguish between discrete and continuous time by using k for discrete and s or t for continuous time.

Define

- the mapping $h : \mathbb{Z}_+^{\mathcal{R}} \times \mathbb{Z}_+^{\mathcal{R}} \mapsto \mathbb{R}$

$$h(y,z) \stackrel{\text{def}}{=} \begin{cases} \log \dfrac{\tilde{\lambda}_r}{\lambda_r} & \text{if } z - y = e_r \text{ and } \tilde{\lambda}_r > 0, \\[2ex] \log \dfrac{\tilde{\mu}_r(y)}{\mu_r(y)} & \text{if } z - y = -e_r \text{ and } \tilde{\mu}_r(y) > 0, \\[2ex] 0 & \text{otherwise.} \end{cases}$$

- the compensator $K : \mathbb{Z}_+^{\mathcal{R}} \mapsto \mathbb{R}$ by

$$\begin{aligned} K(y) &\stackrel{\text{def}}{=} \sum_{z \in \mathbb{Z}_+^{\mathcal{R}}} q(y,z)\Big(e^{h(y,z)} - 1\Big) \\ &= \sum_{r\in\mathcal{R}} \Big(\tilde{\lambda}_r - \lambda_r\Big) + \sum_{r\in\mathcal{R}} \Big(\tilde{\mu}_r(y) - \mu_r(y)\Big). \end{aligned} \tag{5.1}$$

- and the process

$$\mathcal{M}_t \stackrel{\text{def}}{=} \exp\left\{ \sum_{k=0}^{N_t - 1} h\Big(Q(k), Q(k+1)\Big) - \int_0^t K\big(Q(s)\big)\,\mathrm{d}v \right\}.$$

Note that the compensator K is always bounded, so that $\mathcal{M}_t$ takes only finite values. Since K has been exactly defined so that

$$K(x) = \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left[\exp\left\{ \sum_{k=0}^{N_t-1} h\Big(Q(k,x), Q(k+1,x)\Big) \right\} \right]_{t=0},$$

it is easily checked that the derivative of $\mathbb{E}[\mathcal{M}_t]$ at $t = 0$ is null (note that the derivative is independent of Λ, so that it is dropped). Then using the Markov property, one can get that the derivative is null for all $t \geq 0$, so that $\mathbb{E}[\mathcal{M}_t] = 1$. Using again the Markov property, this proves that $\mathbb{E}[\mathcal{M}_t | \mathcal{F}_s] = \mathcal{M}_s$, for all $t \geq s \geq 0$, hence $\{\mathcal{M}_t,\ t \geq 0\}$ is a martingale w.r.t. the natural filtration $\mathcal{F}_t$.

Then define a new probability measure by

$$\tilde{\mathbb{P}}[B] \stackrel{\text{def}}{=} \mathbb{E}\left[\mathbb{1}_{\{B\}}\mathcal{M}_t\right], \quad \forall B \in \mathcal{F}_t.$$

It is a matter of routine to show that under $\tilde{\mathbb{P}}$, Q is again a Markov process. In fact, under $\tilde{\mathbb{P}}$, the system behaves like a localized Markov process where the arrival and the service rates at node r are respectively given by $\tilde{\lambda}_r$ and $\tilde{\mu}_r(y)$ (whence the notation).

Remark : The probability measure $\mathbb{P}$ is not necessarily absolutely continuous with respect to $\tilde{\mathbb{P}}$. This is the case for instance if for some $r \in \mathcal{R}$, $\tilde{\lambda}_r = 0$ (whereas $\lambda_r > 0$).

♦♦♦

5.2 The upper bound

The main step will be to establish the following proposition:

Proposition 5.1 *Let $x \in \mathbb{R}_+^{\mathcal{R}}$ and $G = (A, \eta, D) \in \overline{\mathcal{G}}^x$ be a localized generator. Then*

$$\lim_{\tau,\delta,\epsilon\to 0} \limsup_{n\to\infty} \frac{1}{n\tau} \sup_{|y-nx|<\epsilon n} \log \mathbb{P}\left[E_{\tau,\delta,y}^{(n)}(x,G)\right] \leq -H\big(G\|R(x)\big)$$

where $E_{\tau,\delta,y}^{(n)}(x,G)$ is the event defined in (3.1). Moreover, if a face Λ and a growth rate $D \in \mathbb{R}^\Lambda$ are fixed, then the preceding limit in τ is uniform w.r.t. x in compact sets of Λ (see Definition 1.3).

Proof : Take $G \in \overline{\mathcal{G}}^x$. It is associated to a multi-dimensional birth and death process $(\tilde{\lambda}_r, \tilde{\mu}_r \nu_r(y),\ y \in \mathbb{R}_+^{\mathcal{R}}, r \in \mathcal{R})$ with law $\tilde{\mathbb{P}}$ where the parameters $\tilde{\lambda}_r$, $\tilde{\mu}_r$ are specified according to (3.9). Since $\mathbb{P}$ is not necessarily absolutely continuous w.r.t. $\tilde{\mathbb{P}}$, we introduce a sequence of multi-dimensional birth and death processes $(\tilde{\lambda}_r^{(\kappa)}, \tilde{\mu}_r^{(\kappa)} \nu_r(y), y \in \mathbb{R}_+^{\mathcal{R}}, r \in \mathcal{R})$ with law $\tilde{\mathbb{P}}^{(\kappa)}$ such that

$$\begin{aligned}
\tilde{\lambda}_r^{(\kappa)} > 0 \quad \text{and} \quad & \lim_{\kappa\to 0} \tilde{\lambda}_r^{(\kappa)} = \tilde{\lambda}_r, && \forall r \in \mathcal{R}, \\
\tilde{\mu}_r^{(\kappa)} > 0 \quad \text{and} \quad & \lim_{\kappa\to 0} \tilde{\mu}_r^{(\kappa)} = \tilde{\mu}_r, && \forall r \in \Lambda(x), \\
& \tilde{\mu}_r^{(\kappa)} = \tilde{\mu}_r && \forall r \in \Lambda^c(x).
\end{aligned}$$

In this setting, $\{\mathcal{M}_t^{(\kappa)},\ t \geq 0\}$ is the martingale defining $\tilde{\mathbb{P}}^{(\kappa)}$ with respect to $\mathbb{P}$ and $h^{(\kappa)}(x,y)$ and $K^{(\kappa)}(x)$ are the functions used to defined $\mathcal{M}_t^{(\kappa)}$ according to Section 5.1. Now, $\tilde{\mathbb{P}}^{(\kappa)}$ and $\mathbb{P}$ are mutually absolutely continuous and,

$$\mathbb{P}\left[E_{\tau,\delta,y}^{(n)}(x,G)\right] = \tilde{\mathbb{E}}^{(\kappa)}\left[\mathbb{1}_{\{E_{\tau,\delta,y}^{(n)}(x,G)\}} \left(\mathcal{M}_{n\tau}^{(\kappa)}\right)^{-1}\right]. \tag{5.2}$$

Let us majorize $\left(\mathcal{M}_{n\tau}^{(\kappa)}\right)^{-1}$ on $E_{\tau,\delta,y}^{(n)}(x,G)$ when $|y - nx| < \delta n$. First, recalling from (3.9) that $\tilde{\mu}_r(y) = \mu_r(y)$ for $r \in \Lambda_1(x)$ and $r \in \Lambda_2(x)$ with $\eta_r = 0$ and $y \in \mathbb{R}_+^{\mathcal{R}}$, one has the following bounds:

$$
\begin{aligned}
&-\sum_{k=0}^{N_{n\tau}-1} h^{(\kappa)}\big(Q(k),Q(k+1)\big) \qquad (5.3)\\
&\leq \; -n\tau\Bigg(\sum_{r\in\mathcal{R}}\tilde\lambda_r^{(\kappa)}\log\frac{\tilde\lambda_r^{(\kappa)}}{\lambda_r}+\sum_{r\in\Lambda(x)}\tilde\mu_r^{(\kappa)}\nu_r(x)\log\frac{\tilde\mu_r^{(\kappa)}}{\mu_r}\\
&\qquad\qquad+\sum_{r\in\Lambda_2(x)}\mathbb{I}_{\{\eta_r>0\}}\tilde\mu_r^{(\kappa)}\eta_r\log\frac{\tilde\mu_r^{(\kappa)}}{\mu_r}\Bigg)\\
&\quad+n\tau\delta\Bigg(\sum_{r\in\mathcal{R}}\left|\log\frac{\tilde\lambda_r^{(\kappa)}}{\lambda_r}\right|+\sum_{r\in\Lambda(x)}\left|\log\frac{\tilde\mu_r^{(\kappa)}}{\mu_r}\right|+\sum_{r\in\Lambda_2(x)}\mathbb{I}_{\{\eta_r>0\}}\left|\log\frac{\tilde\mu_r^{(\kappa)}}{\mu_r}\right|\Bigg)
\end{aligned}
$$

On $E^{(n)}_{\tau,\delta,y}(x,G)$ the compensator K is bounded in (5.1) by

$$
\begin{aligned}
&\int_0^{n\tau}K^{(\kappa)}\big(Q(s)\big)\,\mathrm{d}s\leq n\tau\sum_{r\in\mathcal{R}}\big(\tilde\lambda_r^{(\kappa)}-\lambda_r\big)\\
&\quad+\; n\tau\sum_{r\in\Lambda(x)}\Bigg((\tilde\mu_r^{(\kappa)}-\mu_r)\nu_r(x)+\sup_{s\in[0,n\tau]}|\nu_r\big(Q(s)\big)-\nu_r(x)||\tilde\mu_r^{(\kappa)}-\mu_r|\Bigg)\\
&\quad+\; n\tau\sum_{r\in\Lambda_2(x)}\mathbb{I}_{\{\eta_r>0\}}\Bigg((\tilde\mu_r^{(\kappa)}-\mu_r)\eta_r+\delta|\tilde\mu_r^{(\kappa)}-\mu_r|\Bigg). \qquad (5.4)
\end{aligned}
$$

On $E^{(n)}_{\tau,\delta,y}(x,G)$, by Assumption (H2), we have for $r\in\Lambda(x)$

$$
\begin{aligned}
0<\nu_r(x) &= \lim_{\tau,\delta,\epsilon\to 0}\liminf_{n\to\infty}\inf_{|y-nx|<\epsilon n}\inf_{s\in[0,n\tau]}\nu_r(Q(s,y)) \qquad (5.5)\\
&= \lim_{\tau,\delta,\epsilon\to 0}\limsup_{n\to\infty}\sup_{|y-nx|<\epsilon n}\sup_{s\in[0,n\tau]}\nu_r(Q(s,y)).
\end{aligned}
$$

Finally, bounding $\mathbb{I}_{\{E^{(n)}_{\tau,\delta,y}(x,G)\}}$ by 1, bounding $\mathcal{M}^{(\kappa)}_{n\tau}$ using (5.3), (5.4) and (5.5) and taking into account the order in which the different limits are taken, the representation formula (5.2) yields

$$\begin{aligned}
&\lim_{\tau,\delta,\epsilon\to 0} \limsup_{n\to\infty} \frac{1}{n\tau} \sup_{|y-nx|<\epsilon n} \log \mathbb{P}\left[E^{(n)}_{\tau,\delta,y}(x,G)\right] \\
\le\ & -\sum_{r\in\mathcal{R}} \left(\tilde{\lambda}_r \log \frac{\tilde{\lambda}^{(\kappa)}_r}{\lambda_r} - \tilde{\lambda}^{(\kappa)}_r + \lambda_r\right) \\
& - \sum_{r\in\Lambda(x)} \nu_r(x) \left(\tilde{\mu}^{(\kappa)}_r \log \frac{\tilde{\mu}^{(\kappa)}_r}{\mu_r} - \tilde{\mu}^{(\kappa)}_r + \mu_r\right) \\
& - \sum_{r\in\Lambda_2(x)} \mathbb{I}_{\{\eta_r>0\}} \eta_r \left(\tilde{\mu}^{(\kappa)}_r \log \frac{\tilde{\mu}^{(\kappa)}_r}{\mu_r} - \tilde{\mu}^{(\kappa)}_r + \mu_r\right).
\end{aligned}$$

The proof of the upper bound is concluded letting κ tends to 0. ■

5.3 The lower bound

First, we simplify the bound of Theorem 1.7.

Lemma 5.2 *Let $x \in \mathbb{R}^{\mathcal{R}}_+$ and $D \in \mathbb{R}^{\Lambda(x)}$. Then,*

$$\begin{aligned}
&\lim_{\tau,\delta,\epsilon\to 0} \liminf_{n\to\infty} \frac{1}{n\tau} \inf_{|y-nx|<\epsilon n} \log \mathbb{P}\left[\sup_{t\in[0,n\tau]} |Q(t,y) - nx - Dt| < \delta n\right] \\
=\ & \lim_{\tau,\delta\to 0} \liminf_{n\to\infty} \frac{1}{n\tau} \log \mathbb{P}\left[\sup_{t\in[0,n\tau]} |Q(t,\lfloor nx\rfloor) - nx - Dt| < \delta n\right].
\end{aligned}$$

Moreover, if a face Λ and a growth rate $D \in \mathbb{R}^{\Lambda}$ are fixed, then the preceding limit in τ is uniform w.r.t. x in compact sets of Λ (see Definition 1.3).

Proof : The proof of this lemma is postponed to Appendix B. ■

Then, the main step in the proof of the lower bound is to establish

Proposition 5.3 *Let $x \in \mathbb{R}^{\mathcal{R}}_+$ and $G = (A,\eta,D) \in \overline{\mathcal{G}}^x$ be a localized generator. Then*

$$\lim_{\tau,\delta\to 0} \liminf_{n\to\infty} \frac{1}{n\tau} \log \mathbb{P}\left[E^{(n)}_{\tau,\delta,x}(x,G)\right] \ge -H\big(G\|R(x)\big)$$

where $E^{(n)}_{\tau,\delta,y}(x,G)$ is the event defined in (3.1). Moreover, if a face Λ and a growth rate $D \in \mathbb{R}^{\Lambda}$ are fixed, then the preceding limit in τ is uniform w.r.t. x in compact sets of Λ (see Definition 1.3).

Proof : First, the proof is establish for $G \in \mathcal{G}^x$. To G is associated a localized Markov process $(\tilde\lambda_r, \tilde\mu_r \nu_r^x(y), y \in \mathbb{R}_+^{\mathcal{R}}, r \in \mathcal{R})$ with law $\tilde{\mathbb{P}}$ where the parameters $\tilde\lambda_r$, $\tilde\mu_r$ are specified according to (3.9). Denote the event (appearing in Lemma 3.4)

$$F^{(n)}_{\tau,\delta,y}(x,G) \stackrel{\text{def}}{=} E^{(n)}_{\tau,y}(x,G) \cap \big\{A_r(n\tau)=0,\ \forall r \ with\ a_r=0\big\}.$$

Although $\mathbb{P}$ is not absolutely continuous w.r.t $\tilde{\mathbb{P}}$, by definition of $\mathcal{G}^x$, $\tilde\lambda_r > 0$ and $\tilde\mu_r > 0$, $\forall r \in \Lambda(x)$ so that $\mathbb{P}$ is absolutely continuous w.r.t $\tilde{\mathbb{P}}$ on $F^{(n)}_{\tau,\delta,y}(x,D)$ and

$$\begin{aligned}
\mathbb{P}\left[E^{(n)}_{\tau,\delta,x}(x,G)\right] &\geq \mathbb{P}\left[F^{(n)}_{\tau,\delta,x}(x,G)\right] \\
&\geq \inf_{\omega \in F^{(n)}_{\tau,\delta,x}(x,D)} \Big(\mathcal{M}_{n\tau}(\omega)\Big)^{-1} \tilde{\mathbb{P}}\left[F^{(n)}_{\tau,\delta,x}(x,G)\right].
\end{aligned}$$

Let us minorize $\Big(\mathcal{M}_{n\tau}(\omega)\Big)^{-1}$ on $F^{(n)}_{\tau,\delta,x}(x,G)$. Since there is no departures from routes belonging to $\Lambda_1(x)$ and for the routes $r \in \Lambda_2(x)$ with $\eta_r = 0$, we get

$$\begin{aligned}
&-\sum_{k=0}^{N_{n\tau}-1} h^{(\kappa)}\big(Q(k),Q(k+1)\big) \qquad (5.6)\\
\geq\ & -n\tau \sum_{r\in\mathcal{R}} \tilde\lambda_r \log\frac{\tilde\lambda_r}{\lambda_r} + n\tau \sum_{r\in\Lambda(x)} \tilde\mu_r\nu_r(x) \log \sup_{s\in[0,n\tau]} \frac{\tilde\mu_r\nu_r(x)}{\mu_r\nu_r\big(Q(s)\big)} \\
& -n\tau \sum_{r\in\Lambda_2(x)} \tilde\mu_r\nu_r(x) \log \sup_{s\in[0,n\tau]} \frac{\tilde\mu_r\nu_r^x\big(Q(s)\big)}{\mu_r\nu_r\big(Q(s)\big)} - n\tau\delta \sum_{r\in\mathcal{R}} \left|\log\frac{\tilde\lambda_r}{\lambda_r}\right| \\
& -n\tau\delta \sum_{r\in\Lambda(x)} \log \sup_{s\in[0,n\tau]} \frac{\tilde\mu_r\nu_r(x)}{\mu_r\nu_r\big(Q(s)\big)} \\
& -n\tau\delta \sum_{r\in\Lambda_2(x)} \mathbb{1}_{\{\eta_r>0\}} \log \sup_{s\in[0,n\tau]} \frac{\tilde\mu_r\nu_r^x\big(Q(s)\big)}{\mu_r\nu_r\big(Q(s)\big)}
\end{aligned}$$

On $F^{(n)}_{\tau,\delta,y}(x,G)$ the compensator K is bounded in (5.1) by

$$\begin{aligned}
\int_0^{n\tau} K^{(\kappa)}\big(Q(s)\big)\,\mathrm{d}s \geq n\tau \sum_{r\in\mathcal{R}} \big(\tilde\lambda_r - \lambda_r\big) \qquad (5.7)\\
+ \quad n\tau \sum_{r\in\Lambda(x)} \Bigg((\tilde\mu_r - \mu_r)\nu_r(x) - \sup_{s\in[0,n\tau]} |\nu_r\big(Q(s)\big) - \nu_r(x)|\mu_r \Bigg)\\
+ \quad n\tau \sum_{r\in\Lambda_2(x)} \mathbb{I}_{\{\eta_r>0\}} \Big((\tilde\mu_r - \mu_r)\eta_r - \delta|\tilde\mu_r - \mu_r| \Big)\\
+ \quad n\tau \sum_{r\in\Lambda_2(x)} \tilde\mu_r \sup_s \Big(\nu_r^x(Q(s)) - \nu_r(Q(s)) \Big).
\end{aligned}$$

Recall that on $F^{(n)}_{\tau,\delta,x}(x,G)$, by assumptions (H2), (H4) we have for $r\in\Lambda(x)$

$$\begin{aligned}
0 < \nu_r(x) &= \lim_{\tau,\delta\to 0} \liminf_{n\to\infty} \inf_{s\in[0,n\tau]} \nu_r(Q(s,x)) \qquad (5.8)\\
&= \lim_{\tau,\delta\to 0} \limsup_{n\to\infty} \sup_{s\in[0,n\tau]} \nu_r(Q(s,x)).
\end{aligned}$$

Moreover, on $F^{(n)}_{\tau,\delta,x}(x,G)$, by assumptions (H3), (H4) we have for $r\in\Lambda_2(x)$

$$\begin{aligned}
0 < 1 &= \lim_{\tau,\delta\to 0} \liminf_{n\to\infty} \inf_{s\in[0,n\tau]} \frac{\nu_r^x(Q(s,x))}{\nu_r(Q(s,x))} \qquad (5.9)\\
&= \lim_{\tau,\delta\to 0} \limsup_{n\to\infty} \sup_{s\in[0,n\tau]} \frac{\nu_r^x(Q(s,x))}{\nu_r(Q(s,x))}.
\end{aligned}$$

By Lemma 3.4, $\tilde{\mathbb{P}}\left[F^{(n)}_{\tau,\delta,x}(x,G)\right]$ tends to 1. Hence, by (5.6), (5.7), (5.8), (5.9),

$$\begin{aligned}
&\lim_{\tau,\delta\to 0} \limsup_{n\to\infty} \frac{1}{n\tau} \log \mathbb{P}\left[F^{(n)}_{\tau,\delta,y}(x,G)\right]\\
\geq & -\sum_{r\in\mathcal{R}} \left(\tilde\lambda_r \log\frac{\tilde\lambda_r}{\lambda_r} - \tilde\lambda_r + \lambda_r \right) - \sum_{r\in\Lambda(x)} \nu_r(x) \left(\tilde\mu_r \log\frac{\tilde\mu_r}{\mu_r} - \tilde\mu_r + \mu_r \right)\\
- & \sum_{r\in\Lambda_2(x)} \mathbb{I}_{\{\eta_r>0\}} \eta_r \left(\tilde\mu_r \log\frac{\tilde\mu_r}{\mu_r} - \tilde\mu_r + \mu_r \right).
\end{aligned}$$

This concludes the proof of the lower bound when $G \in \mathcal{G}^{\Lambda(x)}$. Consider $G \in \overline{\mathcal{G}}^x$. For any $\delta > 0$, there exists $G' \in \mathcal{G}^x$ and $\delta' > 0$ such that $B(G', \delta') \subset B(G, \delta)$. Hence

$$\begin{aligned} &\lim_{\tau,\delta\to 0} \liminf_{n\to\infty} \frac{1}{n\tau} \log \mathbb{P}\left[E^{(n)}_{\tau,\delta,x}(x,G)\right] \\ \geq \quad &\lim_{\tau,\delta'\to 0} \liminf_{n\to\infty} \frac{1}{n\tau} \log \mathbb{P}\left[E^{(n)}_{\tau,\delta',x}(x,G')\right] = -H(G'\|R(x)). \end{aligned}$$

Since this is true for any $G' \in \mathcal{G}^{\Lambda(x)}$ arbitrary closed to G, by continuity of the entropy (see Lemma 4.2), the lower bound is proved for any $G \in \overline{\mathcal{G}}^{\Lambda(x)}$.

■

6 Sample path LDP

This section aims at proving how to derive Theorem 1.7 and the sample path LDP (Theorem 1.2) from local bounds for empirical generators (Propositions 5.1 and 5.3). Since there is nothing new in this derivation, proof are sketched and references are given for more details.

6.1 Exponential tightness and contraction

Propositions 5.1 and 5.3 imply Theorem 1.7. This step is described in [4, Sections 4.3 and 4.4].

In order to prove this, it is first necessary to check for the exponential tightness of the sequence of probabilities $\mathbb{P}[G_t \in .]$. Actually, the first two components (i.e. A and η) of the empirical generator evolve in compact sets; the last component (i.e. the growth rate D) evolves in $\mathbb{R}^{\mathcal{R}}$ but it is easily checked that the exponential decay of the probability for one coordinate to be larger than τ tends to infinity with τ since all jump rates are bounded (so that this probability is bounded by a Poisson process, which is exponentially tight).

Then the continuity of the mapping $f_x(G) = D$ is used to obtains Theorem 1.7 by a kind of contraction principle (which requires the exponential tightness). This explains the form of $L(x, D)$ in equation (4.2), which is the classical form of the rate function obtained by contraction.

6.2 Large deviations for linear paths

Theorem 1.7 gives bounds only for infinitesimally small linear pieces (when $\tau \to 0$). The first step to get the sample path LDP Theorem 1.2 is to calculate large deviations

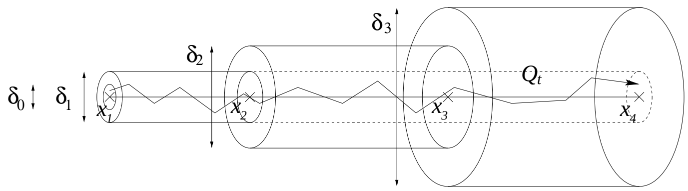


Figure 4: Gluing small linear pieces.

bounds for linear paths. This is done by splitting (using the Markov property) a linear path $[x, x+DT]$ into K small pieces $[x+iDT/K, x+(i+1)DT/K]$ with corresponding δ_i and ε_i verifying $\varepsilon_{i+1} > \delta_i$ as represented in Figure 4. The rate function for such linear paths is precisely I_T defined in equation (1.2), with the simple expression

$$I_T(\phi) \stackrel{\text{def}}{=} \int_0^T L\,(x+Dt, D)\,dt, \qquad \forall \phi:\; t \to x+Dt.$$

6.3 Boundary effects

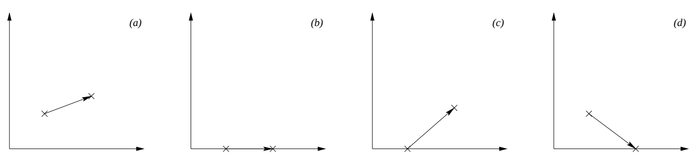


Figure 5: Extremal parts of linear paths.

There is still a difficulty when "gluing" infinitesimally small linear pieces as in the previous section: we did not took into account possible boundary effects occurring at the extremal parts of linear paths. Indeed, if we take a look at Figure 5, it appears that a linear path can change of face at these points (cases (c) and (d)), though Theorem 1.7 deals only with small linear paths staying *inside* a face.

This matter is treated in [5, Appendix B]. The answer is that case (c) does not offer any difficulty: set the arrival rates to the right values (i.e. to get growth rate D); the large deviation cost for that is finite but the length of path tends to 0, hence there is no extra cost for this extremal part.

Case (d) is more difficult: departure rate can become null when approaching the boundary thus the cost to get the right D becomes infinite. There it is necessary to come into more details. This is where Assumption (H6) is necessary (compare with [5, equation (B.3)]): if the departure rate decrease to 0 not too quickly[2], a fine analysis shows there is no extra cost in this case either.

6.4 Absolutely continuous paths

The step from linear paths to piecewise linear paths is exactly the same as for linear paths: glue pieces together using the Markov property. The rate function is still I_T.

The bounds are extended to absolutely continuous paths with finite rate using approximations (see remark p.25) and the properties of I_T (see Proposition 4.6). The reader is referred to [7, Section 5] for details.

6.5 Sample path LDP

The lower bound is immediately proved by using the lower bound for absolutely continuous paths with finite rate. I_T is a good rate function by Proposition 4.6.

The upper bound is more involved: the set of absolutely continuous paths is dense in $D\big([0,T],\mathbb{R}_+^{\mathcal{R}}\big)$, but this does not mean we can cover $D\big([0,T],\mathbb{R}_+^{\mathcal{R}}\big)$ with the neighborhoods of each ones. Hence we use a completely different approach to prove the upper bound: this is Theorem 4.7. The identification thereafter (Proposition 4.8) shows that the upper and lower bounds coincide: the sample path LDP is proved.

Appendix A Proof of Lemma 3.3

There are two cases to consider, associated to equations (3.4) and (3.5). The guideline of the proof is "when the services are asymptotically cut due to scaling, the cost to turn the service rate to a given strictly positive rate is infinite at a large deviation scale".

Case 1: $a_{r_1} > 0$ for some $r_1 \in \Lambda_1(x)$.

The proof uses a change of measure as described in Section 5.1, which we keep the notation of. This change of measure is defined by modifying the intensities of jump

[2] A careful inspection of the proof shows that any polynomial rate is "not too quick" (e.g. $x_r^2/\|x\|$), but we stay to the bound of Assumption (H6) for simplicity. Moreover it is sufficient in most applications, as shown in Section 2.

only on route r_1:

$$\tilde{\lambda}_{r_1} = a_{r_1},$$
$$\tilde{\mu}_{r_1}(y) = \tilde{\mu}_{r_1}, \qquad \forall y \in \mathbb{R}_+^{\mathcal{R}},$$

where $\tilde{\mu}_{r_1} > 0$ is an arbitrary number. Then $\mathbb{P}$ and $\tilde{\mathbb{P}}$ are mutually absolutely continuous and

$$\mathbb{P}\left[E^{(n)}_{\tau,\delta,y}(x,G)\right] = \tilde{\mathbb{E}}\left[\mathbb{1}_{\{E^{(n)}_{\tau,\delta,y}(x,G)\}}\left(\mathcal{M}_{n\tau}\right)^{-1}\right]. \tag{A.1}$$

Let us majorize $\mathcal{M}_{n\tau}^{-1}$ more explicitly on $E^{(n)}_{\tau,\delta,y}(x,G)$. First, using the fact that $G_t \in B(G,\delta)$, the number of arrivals (or departures) from route r_1 is less than $n\tau(\tilde{\lambda}_{r_1}+\delta)$ so that

$$\begin{aligned}
&-\sum_{k=0}^{N_{n\tau}-1} h\big(Q(k),Q(k+1)\big)\\
\leq\ & n\tau(\tilde{\lambda}_{r_1}+\delta)\left(\left|\log\frac{\tilde{\lambda}_{r_1}}{\lambda_{r_1}}\right| - \inf_{z\in B_{\tau,\delta}}\log\frac{\tilde{\mu}_{r_1}}{\mu_{r_1}\nu_{r_1}(nz)}\right)
\end{aligned}$$

where

$$B_{\tau,\delta} = \{z : z \in B(x+Dt,\delta) \text{ for some } t\in[0,\tau]\}\,.$$

Second, all quantities in the compensator $K(.)$ (see equation 5.1) are bounded w.r.t. x. So there exists a constant K_+ such that $K(x) \leq K_+$ for all $x \in \mathbb{R}_+^{\mathcal{R}}$ and we get the bound

$$\frac{1}{n\tau}\log\mathcal{M}_{n\tau}^{-1} \leq K_+ + (\tilde{\lambda}_{r_1}+\delta)\left(\left|\log\frac{\tilde{\lambda}_{r_1}}{\lambda_{r_1}}\right| - \inf_{z\in B_{\tau,\delta}}\log\frac{\tilde{\mu}_{r_1}}{\mu_{r_1}\nu_{r_1}(nz)}\right)$$

Combining this bound with (A.1) yields

$$\begin{aligned}
&\lim_{\tau,\delta,\epsilon\to 0}\limsup_{n\to\infty}\frac{1}{n\tau}\sup_{|y-nx|<n\epsilon}\log\mathbb{P}\left[E^{(n)}_{\tau,\delta,y}(x,D)\right]\\
\leq\ & K_+ + \tilde{\lambda}_{r_1}\left|\log\frac{\tilde{\lambda}_{r_1}}{\lambda_{r_1}}\right| - \tilde{\lambda}_{r_1}\lim_{\tau\to 0}\lim_{n\to\infty}\inf_{z\in B_{\tau,\delta}}\log\frac{\tilde{\mu}_{r_1}}{\mu_{r_1}\nu_{r_1}(nz)} = -\infty.
\end{aligned}$$

Indeed, $r_1 \in \Lambda_1(x)$ so that $\nu^x_{r_1}(.)$ is null and using (H3),

$$\lim_{\tau\to 0}\lim_{n\to\infty}\sup_{z\in B_{\tau,\delta}}\nu_{r_1}(nz) = 0.$$

Now, since $\tilde{\lambda}_{r_1} > 0$ and $\tilde{\mu}_{r_1} > 0$, the limit is infinite.

Case 2: For some $r_1 \in \Lambda_2(x)$, $\eta_{r_1} = 0$ and $a_{r_1} > 0$.

The proof uses a change of measure. The intensities of jump are modified only on route r_1. Take

$$\tilde{\lambda}_{r_1} = a_{r_1},$$
$$\tilde{\mu}_{r_1}(y) = \tilde{\mu}_{r_1}\nu_{r_1}(y), \qquad \forall y \in \mathbb{R}_+^{\mathcal{R}},$$

where $\tilde{\mu}_{r_1} > 0$ is an arbitrary number. Then $\mathbb{P}$ and $\tilde{\mathbb{P}}$ are mutually absolutely continuous. Using $\eta_{r_1} = 0$,

$$\begin{aligned}
&\lim_{\tau,\delta,\epsilon\to 0} \limsup_{n\to\infty} \frac{1}{n\tau} \sup_{|y-nx|<n\epsilon} \log \mathbb{P}\left[E^{(n)}_{\tau,\delta,y}(x,D)\right] \\
&\leq \lim_{\tau,\delta,\epsilon\to 0} \limsup_{n\to\infty} \frac{1}{n\tau} \sup_{|y-nx|<n\epsilon} \log \tilde{\mathbb{E}}\left[\mathcal{M}^{-1}_{n\tau}\right] \\
&-\tilde{\lambda}_{r_1}\left(\log\frac{\tilde{\lambda}_{r_1}}{\lambda_{r_1}} + \log\frac{\tilde{\mu}_{r_1}}{\mu_{r_1}}\right) + (\tilde{\lambda}_{r_1} - \lambda_{r_1}).
\end{aligned}$$

The proof is concluded letting $\tilde{\mu}_{r_1}$ tends to $+\infty$. ■

Appendix B Proof of Lemma 5.2

First, it is clear that

$$\begin{aligned}
&\lim_{\tau,\delta,\epsilon\to 0} \liminf_{n\to\infty} \frac{1}{n\tau} \inf_{|y-nx|<\epsilon n} \log \mathbb{P}\left[\sup_{t\in[0,n\tau]} |Q(t,y) - nx - Dt| < \delta n\right] \\
&\leq \lim_{\tau,\delta\to 0} \liminf_{n\to\infty} \frac{1}{n\tau} \log \mathbb{P}\left[\sup_{t\in[0,n\tau]} |Q(t,\lfloor nx\rfloor) - nx - Dt| < \delta n\right].
\end{aligned}$$

Let us prove the opposite inequality. Using Markov property gives

$$\begin{aligned}
&\mathbb{P}\left[\sup_{t\in[0,n\tau]} |Q(t,y) - nx - Dt| < \delta n\right] \\
&\geq \mathbb{P}\left[\sup_{t\in[0,n\epsilon]} |Q(t,y) - nx - Dt| < \delta n, Q(\epsilon n, y) = \lfloor nx\rfloor\right] \\
&* \mathbb{P}\left[\sup_{t\in[0,n(\tau-\epsilon)]} |Q(t,\lfloor nx\rfloor) - nx - D(n\epsilon + t)| < \delta n\right].
\end{aligned}$$

Since

$$\begin{aligned}
&\lim_{\tau,\delta,\epsilon\to 0}\liminf_{n\to\infty}\frac{1}{n\tau}\log\mathbb{P}\left[\sup_{t\in[0,n(\tau-\epsilon)]}|Q(t,\lfloor nx\rfloor)-nx-D(n\epsilon+t)|<\delta n\right]\\
&=\lim_{\tau,\delta\to 0}\liminf_{n\to\infty}\frac{1}{n\tau}\log\mathbb{P}\left[\sup_{t\in[0,n\tau]}|Q(t,\lfloor nx\rfloor)-nx-Dt|<\delta n\right]
\end{aligned}$$

it is enough to prove that

$$\lim_{\tau,\delta,\epsilon\to 0}\liminf_{n\to\infty}\frac{1}{n\tau}\inf_{|y-nx|<\epsilon n}\log\mathbb{P}\left[\sup_{t\in[0,n\epsilon]}|Q(t,y)-nx-Dt|<\delta n, Q(\epsilon n,y)=\lfloor nx\rfloor\right]\geq 0. \tag{B.1}$$

The idea is to exhibit a suitable path. In fact, when $|y-nx|<\epsilon n$, using (H6) and the fact that the intensity of jump is always lower bounded by $\sum_r\lambda_r$ which is strictly positive, it exists C such that

$$\mathbb{P}\left[\sup_{t\in[0,n\epsilon]}|Q(t,y)-nx-Dt|<\delta n, Q(\epsilon n,y)=\lfloor nx\rfloor\right]\geq C\prod_{r\in\mathcal{R}}\left(\prod_{k=0}^{\lfloor n|y-x|\rfloor}\frac{k}{n}\right)$$

at least when δ is sufficiently large w.r.t. ϵ which is not a restriction taking into account the order in which the limits are taken. Hence

$$\begin{aligned}
&\frac{1}{n}\log\mathbb{P}\left[\sup_{t\in[0,n\epsilon]}|Q(t,y)-nx-Dt|<\delta n, Q(\epsilon n,y)=\lfloor nx\rfloor\right]\\
&\geq\ \frac{1}{n}\log C+\sum_{r\in\mathcal{R}}\sum_{k=0}^{\lfloor n|y-x|\rfloor}\frac{1}{n}\log\frac{k}{n}\longrightarrow_{n\to\infty}\sum_{r\in\mathcal{R}}\int_0^{|y-x|}\log x\,dx.
\end{aligned}$$

Taking the different limits, (B.1) is proved and so the proof of the lemma is concluded. ■

References

[1] ATAR, R., AND DUPUIS, P. Large deviations and queueing networks: methods for rate function identification. *Stochastic Process. Appl. 84*, 2 (1999), 255–296.

[2] BERTSEKAS, D., AND GALLAGER, R. *Data networks*. Prentice Hall, 1987.

[3] Bonald, T., and Massoulié, L. Impact of fairness on internet performance. In *Sigmetrics* (2001).

[4] Delcoigne, F., and de La Fortelle, A. Large deviations for polling systems. Tech. Rep. 3892, INRIA, 2000.

[5] Delcoigne, F., and de La Fortelle, A. Large deviations problems for star networks: the min policy. Tech. Rep. 4143, INRIA, 2001.

[6] Dupuis, P., and Ellis, R. S. Large deviations for Markov processes with discontinuous statistics. II. Random walks. *Probab. Theory Related Fields 91*, 2 (1992), 153–194.

[7] Dupuis, P., and Ellis, R. S. The large deviation principle for a general class of queueing systems. I. *Trans. Amer. Math. Soc. 347*, 8 (1995), 2689–2751.

[8] Dupuis, P., Ellis, R. S., and Weiss, A. Large deviations for Markov processes with discontinuous statistics. I. General upper bounds. *Ann. Probab. 19*, 3 (1991), 1280–1297.

[9] Dupuis, P., Ishii, H., and Soner, H. M. A viscosity solution approach to the asymptotic analysis of queueing systems. *Ann. Probab. 18*, 1 (1990), 226–255.

[10] Fayolle, G., la Fortelle, A. D., Lasgouttes, J. M., Massoulié, L., and Roberts, J. Best-effort networks: modeling and performance analysis via large networks asymptotics. In *Proc. of IEEE INFOCOM'01* (2001).

[11] Ignatiouk-Robert, I. Large deviations of Jackson networks. *Ann. Appl. Probab. 10*, 3 (2000), 962–1001.

[12] Ignatiouk-Robert, I. Sample path large deviations and convergence parameters. *Ann. Appl. Probab. 11*, 4 (2001), 1292–1329.

[13] Ignatyuk, I. A., Malyshev, V. A., and Shcherbakov, V. V. The influence of boundaries in problems on large deviations. *Uspekhi Mat. Nauk 49*, 2(296) (1994), 43–102.

[14] Ioffe, A. D., and Tihomirov, V. M. *Theory of extremal problems.* North-Holland Publishing Co., Amsterdam, 1979. Translated from the Russian by Karol Makowski.

[15] MO, J., AND WALRAND, J. Fair end-to-end window-based congestion control. In *IEEE/ACM Trans. on Networking 8-5* (2000), pp. 556–567.

[16] ROCKAFELLAR, R. T. *Convex analysis*. Princeton University Press, Princeton, N.J., 1970. Princeton Mathematical Series, No. 28.

[17] SHWARTZ, A., AND WEISS, A. *Large deviations for performance analysis*. Stochastic Modeling Series. Chapman & Hall, London, 1995. Queues, communications, and computing, With an appendix by Robert J. Vanderbei.

[18] VECIANA, G. D., LEE, T. J., AND KOSTANTOPOULOS, T. Stability and performance analysis of networks supporting services with rate control - could the internet be unstable ? In *Proc. of IEEE INFOCOM'99* (1999).

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38330 Montbonnot-St-Martin (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399